

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C07K 13/00, A61K 39/04 C12N 15/31, G01N 33/569 C12Q 1/68		A1	(11) International Publication Number: WO 91/04272 (43) International Publication Date: 4 April 1991 (04.04.91)
(21) International Application Number: PCT/EP90/01593			(74) Agents: GUTMANN, Ernest et al.; Yves Plasseraud S.A., Boulevard Haussmann 67, F-75008 Paris (FR).
(22) International Filing Date: 19 September 1990 (19.09.90)			(81) Designated States: AU, CA, JP, US.
(30) Priority data: 89402571.7	19 September 1989 (19.09.89)	GB	Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(71) Applicant (for all designated States except US): N.V. IN-NOGENETICS S.A. [BE/BE]; Industriepark Zwijnaarde 7, Box 4, B-9710 Ghent (BE).			
(72) Inventors; and			
(75) Inventors/Applicants (for US only) : CONTENT, Jean [BE/BE]; Avenue Simonne 5, B-1640 Rhode St.-Genèse (BE). DE WIT, Lucas [BE/BE]; Victor Vergauwenstraat 46, B-2670 Puurs (BE). DE BRUYN, Jacqueline [BE/BE]; Hongarijestaat 192, B-1640 Beersel (BE). VAN VOOREN, Jean-Paul [BE/BE]; Brusselbaan 40, B-1600 St-Pieters Leeuw (BE).			

(54) Title: RECOMBINANT POLYPEPTIDES AND PEPTIDES, NUCLEIC ACIDS CODING FOR THE SAME AND USE OF THESE POLYPEPTIDES AND PEPTIDES IN THE DIAGNOSTIC OF TUBERCULOSIS

(57) Abstract

The invention relates to recombinant polypeptides and peptides and particularly to the polypeptide containing in its polypeptidic chain the following amino acid sequence: the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (194) represented in fig. 4a and fig. 4b. The polypeptides and peptides of the invention can be used for the diagnostic of tuberculosis, and can also be part of the active principle in the preparation of vaccine against tuberculosis.

BEST AVAILABLE COPY

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT Austria	ES Spain	MC Monaco
AU Australia	FI Finland	MC Madagascar
BB Barbados	FR France	ML Mali
BE Belgium	GA Gabon	MR Mauritania
BF Burkina Fasso	GB United Kingdom	MW Malawi
BG Bulgaria	CR Greece	NL Netherlands
BJ Benin	HU Hungary	NO Norway
BR Brazil	IT Italy	PL Poland
CA Canada	JP Japan	RO Romania
CF Central African Republic	KP Democratic People's Republic	SD Sudan
CG Congo	of Korea	SE Sweden
CH Switzerland	KR Republic of Korea	SN Senegal
CM Cameroon	LI Liechtenstein	SU Soviet Union
DE Germany	LK Sri Lanka	TO Chad
DK Denmark	LU Luxembourg	TC Togo
		US United States of America

RECOMBINANT POLYPEPTIDES AND PEPTIDES, NUCLEIC ACIDS  
CODING FOR THE SAME AND USE OF THESE POLYPEPTIDES AND  
PEPTIDES IN THE DIAGNOSTIC OF TUBERCULOSIS

---

The invention relates to recombinant polypeptides and peptides, which can be used for the diagnosis of tuberculosis. The invention also relates to a process for preparing the above-said polypeptides and peptides, which are in a state of biological purity such that they can be used as part of the active principle in the preparation of vaccines against tuberculosis.

It also relates to nucleic acids coding for said polypeptides and peptides.

Furthermore, the invention relates to the in vitro diagnostic methods and kits using the above-said polypeptides and peptides and to the vaccines containing the above-said polypeptides and peptides as active principle against tuberculosis.

By "recombinant polypeptides or peptides" it is to be understood that it relates to any molecule having a polypeptidic chain liable to be produced by genetic engineering, through transcription and translation, of a corresponding DNA sequence under the control of appropriate regulation elements within an efficient cellular host. Consequently, the expression "recombinant polypeptides" such as is used herein does not exclude the possibility for the polypeptides to comprise other groups, such as glycosylated groups.

The term "recombinant" indeed involves the fact that the polypeptide has been produced by genetic engineering, particularly because it results from the expression in a cellular host of the corresponding nucleic acid sequences which have previously been introduced into the expression vector used in said host.

Nevertheless, it must be understood that this expression does not exclude the possibility for the polypeptide to be produced by a different process, for instance by classical chemical synthesis according to methods used in the protein synthesis or by proteolytic cleavage of larger molecules.

The expression "biologically pure" or "biological purity" means on the one hand a grade of purity such that the recombinant polypeptide can be used for the production of vaccinating compositions and on the other hand the absence of contaminants, more particularly of natural contaminants.

Tuberculosis remains a major disease in developing countries. The situation is dramatic in some countries, particularly where high incidence of tuberculosis among AIDS patients represents a new source of dissemination of the disease.

Tuberculosis is a chronic infectious disease in which cell-mediated immune mechanisms play an essential role both for protection against and control of the disease.

Despite BCG vaccination, and some effective drugs, tuberculosis remains a major global problem. Skin testing with tuberculin PPD (protein-purified derivative) largely used for screening of the disease is poorly specific, due to cross reactivity with other pathogenic or environmental saprophytic mycobacteria.

Moreover, tuberculin PPD when used in serological tests (ELISA) does not allow to discriminate between patients who have been vaccinated by BCG, or those who have been primo-infected, from those who are developing evolutive tuberculosis and for whom an early and rapid diagnosis would be necessary.

A protein with a molecular weight of 32-kDa has been purified (9) from zinc deficient Mycobacterium bovis BCG culture filtrate (8). This 32-kDa protein of

M. bovis BCG has been purified from Sauton zinc deficient culture filtrate of M. bovis BCG using successively hydrophobic chromatography on Phenyl-Sepharose, ion exchange on DEAE-Sephacel and molecular sieving on Sephadex G-100. The final preparation has been found to be homogeneous as based on several analyses. This  $P_{32}$  protein is a constituent of BCG cells grown in normal conditions. It represents about 3% of the soluble fraction of a cellular extract, and appears as the major protein released in normal Sauton culture filtrate. This protein has been found to have a molecular weight of 32 000 by SDS-polyacrylamide gel electrophoresis and by molecular sieving.

The  $NH_2$ -terminal amino acid sequence of the 32-kDa protein of M. bovis BCG (Phe-Ser-Arg-Pro-Gly-Leu) is identical to that reported for the MPB 59 protein purified from M. bovis BCG substrain Tokyo (34).

Purified  $P_{32}$  of M. bovis BCG has been tested by various cross immunoelectrophoresis techniques, and has been shown to belong to the antigen 85 complex in the reference system for BCG antigens. It has been more precisely identified as antigen 85A in the coss reference system for BCG antigens (7).

Increased levels of immunoglobulin G antibodies towards the 32-kDa protein of M. bovis BCG could be detected in 70% of tuberculous patients (30).

Furthermore, the 32-kDa protein of M. bovis BCG induces specific lymphoproliferation and interferon- $\gamma$  (IFN- $\gamma$ ) production in peripheral blood leucocytes from patients with active tuberculosis (12) and PPD-positive healthy subjects. Recent findings indicate that the amount of 32-kDa protein of M. bovis BCG-induced IFN- $\gamma$  in BCG-sensitized mouse spleen cells is under probable H-2 control (13). Finally, the high affinity of mycobacteria for fibronectin is related to proteins of the BCG 85 antigen complex (1).

Matsuo et al. (17) recently cloned the gene encoding the antigen  $\alpha$ , a major protein secreted by BCG (substrain Tokyo) and highly homologous to MPB 59 antigen in its NH<sub>2</sub>-terminal amino acid sequence, and even identical for its first 6 amino acids : Phe-Ser-Arg-Pro-Gly-Leu.

This gene was cloned by using a nucleotide probe homologous to the N-terminal amino acid sequence of antigen  $\alpha$ , purified from M. tuberculosis as described in Tasaka, H. et al., 1983. "Purification and antigenic specificity of alpha protein (Yoneda and Fukui) from *Mycobacterium tuberculosis* and *Mycobacterium intracellulare*. Hiroshima J. Med. Sci. 32, 1-8.

The presence of antigens of around 30-32-kDa, named antigen 85 complex, has been revealed from electrophoretic patterns of proteins originating from culture media of mycobacteria, such as *Mycobacterium tuberculosis*. By immunoblotting techniques, it has been shown that these antigens cross-react with rabbit sera raised against the 32-kDa protein of BCG (8).

A recent study reported on the preferential humoral response to a 30-kDa and 31-kDa antigen in lepromatous leprosy patients, and to a 32-kDa antigen in tuberculoid leprosy patients (24).

It has also been found that fibronectin (FN)-binding antigens are prominent components of short-term culture supernatants of *Mycobacterium tuberculosis*. In 3-day-old supernatants, a 30-kilodalton (kDa) protein was identified as the major (FN)-binding molecule. In 21-day-old supernatants, FN was bound to a double protein band of around 30 to 32-kDa, as well as to a group of antigens of larger molecular mass (57 to 60 kDa) (1).

In other experiments, recombinant plasmids containing  $\lambda$  DNA from *Mycobacterium tuberculosis* were transformed into *Escherichia coli*, and three colonies

were selected by their reactivity with polyclonal antisera to M. tuberculosis. Each recombinant produced 35- and 53-kilodalton proteins (35K and 53K proteins, respectively) ("Expression of Proteins of *Mycobacterium tuberculosis* in *Escherichia coli* and Potential of Recombinant Genes and Proteins for Development of Diagnostic Reagents", Mitchell L Cohen et al., Journal of Clinical Microbiology, July 1987, p.1176-1180).

Concerning the various results known to date, the physico-chemical characteristics of the antigen P<sub>32</sub> of *Mycobacterium tuberculosis* are not precise and, furthermore, insufficient to enable its unambiguous identifiability, as well as the characterization of its structural and functional elements.

Moreover, the pathogenicity and the potentially infectious property of M. tuberculosis has hampered research enabling to identify, purify and characterize the constituents as well as the secretion products of this bacteria.

An aspect of the invention is to provide recombinant polypeptides which can be used as purified antigens for the detection and control of tuberculosis.

Another aspect of the invention is to provide nucleic acids coding for the peptidic chains of biologically pure recombinant polypeptides which enable their preparation on a large scale.

Another aspect of the invention is to provide antigens which can be used in serological tests as an in vitro rapid diagnostic of tuberculosis.

Another aspect of the invention is to provide a rapid in vitro diagnostic means for tuberculosis, enabling it to discriminate between patients suffering from an evolutive tuberculosis from those who have been vaccinated against BCG or who have been primo-infected.

Another aspect of the invention is to provide nucleic probes which can be used as in vitro diagnostic

reagent for tuberculosis, as well as in vitro diagnostic reagent for identifying M. tuberculosis from other strains of mycobacteria.

The recombinant polypeptides of the invention contain in their polypeptidic chain one at least of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (-29) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b, or
- = the one extending from the extremity constituted by amino acid at position (12) to the extremity constituted by amino acid at position (31) represented on fig. 3a and fig. 3b, or
- the one extending from the extremity constituted by amino acid at position (36) to the extremity constituted by amino acid at position (55) represented on fig. 3a and fig. 3b, or
- the one extending from the extremity constituted by amino acid at position (77) to the extremity constituted by amino acid at position (96) represented on fig. 3a and fig. 3b, or
- the one extending from the extremity constituted by amino acid at position (101) to the extremity constituted by amino acid at position (120) represented on fig. 3a and fig. 3b, or
- the one extending from the extremity constituted by amino acid at position (175) to the extremity constituted by amino acid at position (194) represented on fig. 3a and fig. 3b, or
- the one extending from the extremity constituted by amino acid at position (211) to the extremity constituted by amino acid at position (230) represented on fig. 3a and fig. 3b, or
- the one extending from the extremity constituted by amino acid at position (275) to the extremity

constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,

and the peptidic sequences resulting from the modification by substitution and/or by addition and/or by deletion of one or several amino acids in so far as this modification does not alter the following properties :

the polypeptides react with rabbit polyclonal antiserum raised against the protein of 32-kDa of M. bovis BCG culture filtrate, and/or

react selectively with human sera from tuberculosis patients and particularly patients developing an evolutive tuberculosis at an early stage, and/or react with the amino acid sequence extending from the extremity constituted by amino acid at position (1), to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b.

On figures 3a and 3b :

- X represents G or GG,
- Y represents C or CC,
- Z represents C or G,
- W represents C or G and is different from Z,
- K represents C or CG,
- L represents G or CC,
- a<sub>1</sub>-b<sub>1</sub> represents ALA-ARG or GLY-ALA-ALA,
- a<sub>2</sub> represents arg or gly,
- a<sub>3</sub>-b<sub>3</sub>-c<sub>3</sub>-d<sub>3</sub>-e<sub>3</sub>-f<sub>3</sub>- represents his-trp-val-pro-arg-pro or ala-leu-gly-ala,
- a<sub>4</sub> represents pro or pro-asn-thr,
- a<sub>5</sub> represents pro or ala-pro.

The recombinant polypeptides of the invention contain in their polypeptidic chain one at least of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (-29) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b, or
- the one extending from the extremity constituted by amino acid at position (12) to the extremity constituted by amino acid at position (31) represented on fig. 4a and fig. 4b, or
- the one extending from the extremity constituted by amino acid at position (36) to the extremity constituted by amino acid at position (55) represented on fig. 4a and fig. 4b, or
- the one extending from the extremity constituted by amino acid at position (77) to the extremity constituted by amino acid at position (96) represented on fig. 4a and fig. 4b, or
- the one extending from the extremity constituted by amino acid at position (101) to the extremity constituted by amino acid at position (120) represented on fig. 4a and fig. 4b, or
- the one extending from the extremity constituted by amino acid at position (175) to the extremity constituted by amino acid at position (194) represented on fig. 4a and fig. 4b, or
- the one extending from the extremity constituted by amino acid at position (211) to the extremity constituted by amino acid at position (230) represented on fig. 4a and fig. 4b, or
- the one extending from the extremity constituted by amino acid at position (275) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,

and the peptidic sequences resulting from the modification by substitution and/or by addition and/or by deletion of one or several amino acids in so far as

this modification does not alter the following properties :

the polypeptides react with rabbit polyclonal antiserum raised against the protein of 32-kDa of M. bovis BCG culture filtrate, and/or

react selectively with human sera from tuberculosis patients, and particularly patients developing an evolutive tuberculosis at an early stage, and/or react with the amino acid sequence extending from the extremity constituted by amino acid at position (1), to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b.

The recombinant polypeptides of the invention contain in their polypeptidic chain one at least of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (-30) to the extremity constituted by amino acid at position (-1) represented on fig. 5, or
- the one extending from the extremity constituted by amino acid at position (12) to the extremity constituted by amino acid at position (31) represented on fig. 5, or
- the one extending from the extremity constituted by amino acid at position (36) to the extremity constituted by amino acid at position (55) represented on fig. 5, or
- the one extending from the extremity constituted by amino acid at position (77) to the extremity constituted by amino acid at position (96) represented on fig. 5, or
- the one extending from the extremity constituted by amino acid at position (101) to the extremity constituted by amino acid at position (120) represented on fig. 5, or

- the one extending from the extremity constituted by amino acid at position (175) to the extremity constituted by amino acid at position (194) represented on fig. 5, or

- the one extending from the extremity constituted by amino acid at position (211) to the extremity constituted by amino acid at position (230) represented on fig. 5, or

- the one extending from the extremity constituted by amino acid at position (275) to the extremity constituted by amino acid at position (295) represented on fig. 5,

and the peptidic sequences resulting from the modification by substitution and/or by addition and/or by deletion of one or several amino acids in so far as this modification does not alter the following properties :

the polypeptides react with rabbit polyclonal antiserum raised against the protein of 32-kDa of M. bovis BCG culture filtrate, and/or

react selectively with human sera from tuberculosis patients and particularly patients developing an evolutive tuberculosis at an early stage, and/or react with the amino acid sequence extending from the extremity constituted by amino acid at position (1), to the extremity constituted by amino acid at position (295) represented on fig. 5.

Advantageous polypeptides of the invention are characterized by the fact that they react with rabbit polyclonal antiserum raised against the protein of 32-kDa of M. bovis BCG culture filtrate, hereafter designated by "P<sub>32</sub> protein of BCG".

Advantageous polypeptides of the invention are characterized by the fact that they selectively react with human sera from tuberculous patients and

particularly patients developing an evolutive tuberculosis at an early stage.

Hereafter is given, in a non limitative way a process for preparing rabbit polyclonal antiserum raised against the  $P_{32}$  protein of BCG and a test for giving evidence of the reaction between the polypeptides of the invention and said rabbit polyclonal antiserum raised against the  $P_{32}$  protein of BCG.

1) process for preparing rabbit polyclonal antiserum raised against the  $P_{32}$  protein of BCG:

Purified  $P_{32}$  protein of BCG from culture filtrate is used.

a) Purification of protein  $P_{32}$  of BCG :

$P_{32}$  protein can be purified as follows :

The bacterial strains used are M. bovis BCG substrains 1173P2 (Pasteur Institute, Paris) and GL2 (Pasteur Institute, Brussels).

The culture of bacteria is obtained as follows :

*Mycobacterium bovis* BCG is grown as a pellicle on Sauton medium, at 37.5°C for 14 days. As the medium is prepared with distilled water, zinc sulfate is added to the final concentration of 5  $\mu$ M (normal Sauton medium) (De Bruyn J., Weckx M., Beumer-Jochmans M.-P. Effect of zinc deficiency on *Mycobacterium tuberculosis* var. *bovis* (BCG). J. Gen. Microbiol. 1981; 124:353-7). When zinc deficient medium was needed, zinc sulfate is omitted.

The filtrates from zinc deficient cultures are obtained as follows :

The culture medium is clarified by decantation. The remaining bacteria are removed by filtration through Millipak 100 filter unit (Millipore Corp., Bedford, Mass.). When used for purification, the filtrate is adjusted to 20 mM in phosphate, 450 mM in NaCl, 1 mM in EDTA, and the pH is brought to 7.3 with

5 M HCl before sterile filtration.

The protein analysis is carried out by polyacrylamide gel electrophoresis. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was done on 13% (w/v) acrylamide-containing gels as described by Laemmli UK. (Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 1970; 227:680-5). The gels are stained with Coomassie Brilliant Blue R-250 and for quantitative analysis, scanned at 595 nm with a DU8 Beckman spectrophotometer. For control of purity the gel is revealed with silver stain (Biorad Laboratories, Richmond, Calif.).

The purification step of  $P_{32}$  is carried out as follows:

Except for hydrophobic chromatography on Phenyl-Sepharose, all buffers contain Tween 80 (0.005% final concentration). The pH is adjusted to 7.3 before sterilization. All purification steps are carried out at +4°C. Elutions are followed by recording the absorbance at 280 nm. The fractions containing proteins are analysed by SDS-PAGE.

(i) The treated filtrate from a 4 liters zinc-deficient culture, usually containing 125 to 150 mg protein per liter, is applied to a column (5.0 by 5.0 cm) of Phenyl-Sepharose CL-4B (Pharmacia Fine Chemicals, Uppsala, Sweden), which is previously equilibrated with 20 mM phosphate buffer (PB) containing 0.45 M NaCl and 1 mM EDTA, at a flow rate of 800 ml per hour. The gel is then washed with one column volume of the same buffer to remove unfixed material and successively with 300 ml of 20 mM and 4 mM PB and 10% ethanol (v/v). The  $P_{32}$  appears in the fraction eluted with 10% ethanol.

(ii) After the phosphate concentration of this fraction has been brought to 4 mM, it is applied to a column (2.6 by 10 cm) of DEAE-Sephacel (Pharmacia Fine

Chemicals), which is equilibrated with 4 mM PB. After washing with the equilibrating buffer the sample is eluted with 25 mM phosphate at a flow rate of 50 ml per hour. The eluate is concentrated in a 202 Amicon stirred cell equipped with a PM 10 membrane (Amicon Corp., Lexington, Mass.).

(iii) The concentrated material is submitted to 4 mg of  $P_{32}$  protein of BCG (soluble extract) or molecular sieving on a Sephadex G-100 (Pharmacia) column (2.6 by 45 cm) equilibrated with 50 mM PB, at a flow rate of 12 ml per hour. The fractions of the peak giving one band in SDS-PAGE are pooled. The purity of the final preparation obtained is controlled by SDS-PAGE followed by silverstaining and by molecular sieving on a Superose 12 (Pharmacia) column (12.0 by 30 cm) equilibrated with 50 mM PB containing 0.005% Tween 80 at a flow rate of 0.2 ml/min. in the Fast Protein Liquid Chromatography system (Pharmacia). Elution is followed by recording the absorbance at 280 nm and 214 nm.

b) Preparation of rabbit polyclonal antiserum raised against the  $P_{32}$  protein of BCG :

400  $\mu$ g of purified  $P_{32}$  protein of BCG per ml physiological saline are mixed with one volume of incomplete Freund's adjuvant. The material is homogenized and injected intradermally in 50  $\mu$ l doses delivered at 10 sites in the back of the rabbits, at 0, 4, 7 and 8 weeks (adjuvant is replaced by the diluent for the last injection). One week later, the rabbits are bled and the sera tested for antibody level before being distributed in aliquots and stored at -80°C;

2) test for giving evidence of the reaction between the polypeptides of the invention and said rabbit polyclonal antiserum raised against the  $P_{32}$  protein of BCG:

the test used was an ELISA test; the ELISA for antibody determination is based on the method of Engvall and Perlmann (Engvall, E., and P. Perlmann. 1971. Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. Immunochemistry 8:871-874).

Immuron Microelisa plates (Dynatech, Kloten, Switzerland) are coated by adding to each well 1  $\mu$ g of one of the polypeptides of the invention in 100  $\mu$ l Tris hydrochloride buffer 50 mM (pH 8.2). After incubation for 2 h at 27°C in a moist chamber, the plates are kept overnight at 4°C. They are washed four times with 0.01 M phosphate-buffered saline (pH 7.2) containing 0.05% Tween 20 by using a Titertek microplate washer (Flow Laboratories. Brussels. Belgium). Blocking is done with 0.5% gelatin in 0.06 M carbonate buffer (pH 9.6) for 1 h. Wells are then washed as before, and 100  $\mu$ l of above mentioned serum diluted in phosphate-buffered saline containing 0.05% Tween 20 and 0.5% gelatin is added. According to the results obtained in preliminary experiments, the working dilutions are set at 1:200 for IgG, 1:20 for IgA and 1:80 for IgM determinations. Each dilution is run in duplicate. After 2 h of incubation and after the wells are washed, they are filled with 100  $\mu$ l of peroxidase-conjugated rabbit immunoglobulins directed against human IgG, IgA or IgM (Dakopatts, Copenhagen, Denmark), diluted 1:400, 1:400 and 1:1.200, respectively in phosphate-buffered saline containing 0.05% Tween 20 and 0.5% gelatin and incubated for 90 min. After the wash, the amount of peroxidase bound to the wells is quantified by using a freshly prepared solution of o-phenylenediamine (10 mg/100 ml) and hydrogen peroxide (8  $\mu$ l of 30%  $H_2O_2$  per 100 ml) in 0.15 M citrate buffer (pH 5.0) as a substrate. The enzymatic reaction is stopped with 8 N  $H_2SO_4$  after

15 min. of incubation. The optical density is read at 492 nm with a Titertek Multiskan photometer (Flow Laboratories).

Wells without sera are used as controls for the conjugates. Each experiment is done by including on each plate one negative and two positive reference sera with medium and low antibody levels to correct for plate-to-plate and day-to-day variations. The antibody concentrations are expressed as the optical density values obtained after correction of the readings according to the mean variations of the reference sera.

Hereafter is also given in a non limitative way, a test for giving evidence of the fact that polypeptides of the invention are recognized selectively by human sera from tuberculous patients.

This test is an immunoblotting (Western blotting) analysis, in the case where the polypeptides of the invention are obtained by recombinant techniques. This test can also be used for polypeptides of the invention obtained by a different preparation process. After sodium dodecyl sulfate-polyacrylamide gel electrophoresis, polypeptides of the invention are blotted onto nitrocellulose membranes (Hybond C. (Amersham)) as described by Towbin et al. (29). The expression of polypeptides of the invention fused to  $\beta$ -galactosidase in E. coli Y1089, is visualized by the binding of a polyclonal rabbit anti-32-kDa BCG protein serum (1:1,000) or by using a monoclonal anti- $\beta$ -galactosidase antibody (Promega). The secondary antibody (alkaline phosphatase anti-rabbit immunoglobulin G and anti-mouse alkaline phosphatase immunoglobulin G conjugates, respectively) is diluted as recommended by the supplier (Promega).

In order to identify selective recognition of polypeptides of the invention and of fusion proteins of the invention by human tuberculous sera, nitrocellulose

sheets are incubated overnight with these sera (1:50) (after blocking aspecific protein-binding sites). The human tuberculous sera are selected for their reactivity (high or low) against the purified 32-kDa antigen of BCG tested in a dot blot assay as described in document (31) of the bibliography hereafter. Reactive areas on the nitrocellulose sheets are revealed by incubation with peroxidase conjugated goat anti-human immunoglobulin G antibody (Dakopatts, Copenhagen, Denmark) (1:200) for 4h, and after repeated washings, color reaction is developed by adding peroxidase substrate ( $\alpha$ -chloronaphthol) (Bio-Rad Laboratories, Richmond, Calif.) in the presence of peroxidase and hydrogen peroxide.

It goes without saying that the free reactive functions which are present in some of the amino acids, which are part of the constitution of the polypeptides of the invention, particularly the free carboxyl groups which are carried by the groups Glu or by the C-terminal amino acid on the one hand and/or the free NH<sub>2</sub> groups carried by the N-terminal amino acid or by amino acid inside the peptidic chain, for instance Lys, on the other hand, can be modified in so far as this modification does not alter the above mentioned properties of the polypeptide.

The molecules which are thus modified are naturally part of the invention. The above mentioned carboxyl groups can be acylated or esterified.

Other modifications are also part of the invention. Particularly, the amine or ester functions or both of terminal amino acids can be themselves involved in the bond with other amino acids. For instance, the N-terminal amino acid can be linked to a sequence comprising from 1 to several amino acids corresponding to a part of the C-terminal region of another peptide.

Furthermore, any peptidic sequences resulting from the modification by substitution and/or by addition and/or by deletion of one or several amino acids of the polypeptides according to the invention are part of the invention in so far as this modification does not alter the above mentioned properties of said polypeptides.

The polypeptides according to the invention can be glycosylated or not, particularly in some of their glycosylation sites of the type Asn-X-Ser or Asn-X-Thr, X representing any amino acid.

Advantageous recombinant polypeptides of the invention contain in their polypeptidic chain, one at least of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (-42) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-47) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-49) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-55) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-59) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b.

Advantageous recombinant polypeptides of the invention contain in their polypeptidic chain, one at least of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (-42) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-47) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-49) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-55) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-59) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b.

Advantageous recombinant polypeptides of the invention contain in their polypeptidic chain, one at least of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (-43) to the extremity constituted by amino acid at position (-1) represented on fig. 5.

Advantageous recombinant polypeptides of the invention contain in their polypeptidic chain, one at least of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-29) to the extremity

constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-42) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-47) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-49) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-55) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-59) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b.

Advantageous recombinant polypeptides of the invention contain in their polypeptidic chain, one at least of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,

- the one extending from the extremity constituted by amino acid at position (-29) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,

- the one extending from the extremity constituted by amino acid at position (-42) to the extremity

constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,

- the one extending from the extremity constituted by amino acid at position (-47) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,

- the one extending from the extremity constituted by amino acid at position (-49) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,

- the one extending from the extremity constituted by amino acid at position (-55) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,

- the one extending from the extremity constituted by amino acid at position (-59) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b.

Advantageous recombinant polypeptides of the invention contain in their polypeptidic chain, one at least of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (295) represented on fig. 5,

- the one extending from the extremity constituted by amino acid at position (-30) to the extremity constituted by amino acid at position (295) represented on fig. 5,

- the one extending from the extremity constituted by amino acid at position (-43) to the extremity constituted by amino acid at position (295) represented on fig. 5.

Other advantageous recombinant polypeptides of the invention consist in one of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (-59) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-55) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-49) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-47) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-42) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-29) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b.

Other advantageous recombinant polypeptides of the invention consist in one of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (-59) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,

- the one extending from the extremity constituted by amino acid at position (-55) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-49) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-47) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-42) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-29) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b.

Other advantageous recombinant polypeptides of the invention consist in one of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (295) represented on fig. 5,
- the one extending from the extremity constituted by amino acid at position (-30) to the extremity constituted by amino acid at position (295) represented on fig. 5,
- the one extending from the extremity constituted by amino acid at position (-43) to the extremity

constituted by amino acid at position (295) represented on fig. 5.

Other advantageous recombinant polypeptides of the invention consist in one of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (-59) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-55) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-49) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-47) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-42) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-29) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b.

Other advantageous recombinant polypeptides of the invention consist in one of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (-59) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,

- the one extending from the extremity constituted by amino acid at position (-55) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-49) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-47) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-42) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-29) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b.

Other advantageous recombinant polypeptides of the invention consist in one of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (-43) to the extremity constituted by amino acid at position (-1) represented on fig. 5,
- the one extending from the extremity constituted by amino acid at position (-30) to the extremity constituted by amino acid at position (-1) represented on fig. 5.

In eukaryotic cells, these polypeptides can be used as signal peptides, the role of which is to initiate the translocation of a protein from its site of synthesis, but which is excised during translocation.

Other advantageous peptides of the invention consist in one of the following amino acid sequence:

- the one extending from the extremity constituted by amino acid at position (12) to the extremity constituted by amino acid at position (31) represented on fig. 3a and fig. 3b, or
- the one extending from the extremity constituted by amino acid at position (36) to the extremity constituted by amino acid at position (55) represented on fig. 3a and fig. 3b, or
- the one extending from the extremity constituted by amino acid at position (77) to the extremity constituted by amino acid at position (96) represented on fig. 3a and fig. 3b, or
- the one extending from the extremity constituted by amino acid at position (101) to the extremity constituted by amino acid at position (120) represented on fig. 3a and fig. 3b, or
- the one extending from the extremity constituted by amino acid at position (175) to the extremity constituted by amino acid at position (194) represented on fig. 3a and fig. 3b, or
- the one extending from the extremity constituted by amino acid at position (211) to the extremity constituted by amino acid at position (230) represented on fig. 3a and fig. 3b, or
- the one extending from the extremity constituted by amino acid at position (275) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b.

Other advantageous peptides of the invention consist in one of the following amino acid sequence:

- the one extending from the extremity constituted by amino acid at position (12) to the extremity constituted by amino acid at position (31) represented on fig. 4a and fig. 4b, or

- the one extending from the extremity constituted by amino acid at position (36) to the extremity constituted by amino acid at position (55) represented on fig. 4a and fig. 4b, or
- the one extending from the extremity constituted by amino acid at position (77) to the extremity constituted by amino acid at position (96) represented on fig. 4a and fig. 4b, or
- the one extending from the extremity constituted by amino acid at position (101) to the extremity constituted by amino acid at position (120) represented on fig. 4a and fig. 4b, or
- the one extending from the extremity constituted by amino acid at position (175) to the extremity constituted by amino acid at position (194) represented on fig. 4a and fig. 4b, or
- the one extending from the extremity constituted by amino acid at position (211) to the extremity constituted by amino acid at position (230) represented on fig. 4a and fig. 4b, or
- the one extending from the extremity constituted by amino acid at position (275) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b.

Other advantageous peptides of the invention consist in one of the following amino acid sequence:

- the one extending from the extremity constituted by amino acid at position (12) to the extremity constituted by amino acid at position (31) represented on fig. 5, or
- the one extending from the extremity constituted by amino acid at position (36) to the extremity constituted by amino acid at position (55) represented on fig. 5, or
- the one extending from the extremity constituted by amino acid at position (77) to the extremity

constituted by amino acid at position (96) represented on fig. 5, or

- the one extending from the extremity constituted by amino acid at position (101) to the extremity constituted by amino acid at position (120) represented on fig. 5, or

- the one extending from the extremity constituted by amino acid at position (175) to the extremity constituted by amino acid at position (194) represented on fig. 5, or

- the one extending from the extremity constituted by amino acid at position (211) to the extremity constituted by amino acid at position (230) represented on fig. 5, or

- the one extending from the extremity constituted by amino acid at position (275) to the extremity constituted by amino acid at position (295) represented on fig. 5.

It is to be noted that the above mentioned polypeptides are derived from the expression products of a DNA derived from the nucleotide sequence coding for a protein of 32-kDa secreted by *Mycobacterium tuberculosis* as explained hereafter in the examples.

The invention also relates to the amino acid sequences constituted by the above mentioned polypeptides and a protein or an heterologous sequence with respect to said polypeptide, said protein or heterologous sequence comprising for instance from about 1 to about 1000 amino acids. These amino acid sequences will be called fusion proteins.

In an advantageous fusion protein of the invention, the heterologous protein is  $\beta$ -galactosidase.

Other advantageous fusion proteins of the invention are the ones containing an heterologous protein resulting from the expression of one of the following plasmids:

pEX1  
pEX2  
pEX3  
pUEX1  
pUEX2  
pUEX3

The invention also relates to any nucleotide sequence coding for a polypeptide of the invention.

The invention also relates to nucleic acids comprising nucleotide sequences which hybridize with the nucleotide sequences coding for any of the above mentioned polypeptides under the following hybridization conditions:

- hybridization and wash medium: 3 X SSC, 20% formamide (1 X SSC is 0,15 M NaCl, 0.015 M sodium citrate, pH 7.0),

- hybridization temperature (HT) and wash temperature (WT) for the nucleic acids of the invention defined by x-y: i.e. by the sequence extending from the extremity constituted by the nucleotide at position (x) to the extremity constituted by the nucleotide at position (y) represented on fig. 3a and fig. 3b.

1 - 182	HT = WT = 69°C
1 - 194	HT = WT = 69°C
1 - 212	HT = WT = 69°C
1 - 218	HT = WT = 69°C
1 - 272	HT = WT = 69°C
1 - 359	HT = WT = 71°C
1 - 1241	HT = WT = 73°C
1 - 1358	HT = WT = 73°C
183 - 359	HT = WT = 70°C
183 - 1241	HT = WT = 73°C
183 - 1358	HT = WT = 73°C
195 - 359	HT = WT = 70°C
195 - 1241	HT = WT = 73°C
195 - 1358	HT = WT = 73°C
213 - 359	HT = WT = 70°C

213 - 1241	HT = WT = 73°C
213 - 1358	HT = WT = 73°C
219 - 359	HT = WT = 71°C
219 - 1241	HT = WT = 73°C
219 - 1358	HT = WT = 73°C
234 - 359	HT = WT = 71°C
234 - 1241	HT = WT = 74°C
234 - 1358	HT = WT = 73°C
273 - 359	HT = WT = 71°C
273 - 1241	HT = WT = 74°C
273 - 1358	HT = WT = 73°C
360 - 1241	HT = WT = 73°C
360 - 1358	HT = WT = 73°C
1242 - 1358	HT = WT = 62°C

The above mentioned temperatures are to be considered as approximately  $\pm 5^\circ\text{C}$ .

The invention also relates to nucleic acids comprising nucleotide sequences which are complementary to the nucleotide sequences coding for any of the above mentioned polypeptides.

It is to be noted that in the above defined nucleic acids, as well as in the hereafter defined nucleic acids, the nucleotide sequences which are brought into play are such that T can be replaced by U.

A group of preferred nucleic acids of the invention comprises one at least of the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (182) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (360) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (1242) to the extremity constituted by nucleotide at position (1358), wherein N represents one of the five A, T, C, G or I nucleotides, represented in fig. 3a and fig. 3b,

or above said nucleotide sequences wherein T is replaced by U,

or nucleic acids which hybridize with said above mentioned nucleotide sequences or the complementary sequences thereof.

A group of preferred nucleic acids of the invention comprises one at least of the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (182) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (360) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (1242) to the extremity constituted by nucleotide at position (1358), wherein N represents one of the five A, T, C, G or I nucleotides, represented in fig. 4a and fig. 4b,

or above said nucleotide sequences wherein T is replaced by U,

or nucleic acids which hybridize with said above mentioned nucleotide sequences or the complementary sequences thereof.

A group of preferred nucleic acids of the invention comprises one at least of the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (130) to the extremity constituted by nucleotide at position (219) represented in fig. 5,

- the one extending from the extremity constituted by nucleotide at position (220) to the extremity constituted by nucleotide at position (1104) represented in fig. 5,

- the one extending from the extremity constituted by nucleotide at position (1104) to the extremity constituted by nucleotide at position (1299), wherein N represents one of the five A, T, C, G or I nucleotides, represented in fig. 5,

or above said nucleotide sequences wherein T is replaced by U,

or nucleic acids which hybridize with said above mentioned nucleotide sequences or the complementary sequences thereof.

Other preferred nucleic acids of the invention comprise one at least of the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b.

Other preferred nucleic acids of the invention comprise one at least of the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b.

Another preferred group of nucleic acids of the invention comprises the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (360) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b.

Another preferred group of nucleic acids of the invention comprises the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (360) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b.

According to another advantageous embodiment, nucleic acids of the invention comprises one of the following sequences:

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (194) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (212) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (218) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (272) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b.

According to another advantageous embodiment, nucleic acids of the invention comprises one of the following sequences:

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (194) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (212) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (218) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (272) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b.

Preferred nucleic acids of the invention consist in one of the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (219) to the extremity

constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b.

Preferred nucleic acids of the invention consist in one of the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b.

These nucleotide sequence can be used as nucleotide signal sequences, coding for the corresponding signal peptide.

Preferred nucleic acids of the invention consist in one of the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (360) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (360) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b.

Preferred nucleic acids of the invention consist in one of the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (182) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (194) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (212) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (218) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (272) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1242) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b.

Preferred nucleic acids of the invention consist in one of the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (360) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (360) to the extremity

constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b.

Preferred nucleic acids of the invention consist in one of the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (182) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (194) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (212) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (218) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (272) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1242) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b.

Preferred nucleic acids of the invention consist in one of the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (129) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (219) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1104) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1299) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (90) to the extremity constituted by nucleotide at position (219) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (90) to the extremity

constituted by nucleotide at position (1299) represented in fig. 5,

- the one extending from the extremity constituted by nucleotide at position (90) to the extremity constituted by nucleotide at position (1104) represented in fig. 5,

- the one extending from the extremity constituted by nucleotide at position (130) to the extremity constituted by nucleotide at position (1104) represented in fig. 5,

- the one extending from the extremity constituted by nucleotide at position (130) to the extremity constituted by nucleotide at position (1299) represented in fig. 5,

- the one extending from the extremity constituted by nucleotide at position (220) to the extremity constituted by nucleotide at position (1299) represented in fig. 5.

Preferred nucleic acids of the invention consist in one of the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (129) represented in fig. 5,

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (219) represented in fig. 5,

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1104) represented in fig. 5,

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1299) represented in fig. 5,

- the one extending from the extremity constituted by nucleotide at position (90) to the extremity constituted by nucleotide at position (219) represented in fig. 5,

- the one extending from the extremity constituted by nucleotide at position (90) to the extremity constituted by nucleotide at position (1104) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (90) to the extremity constituted by nucleotide at position (1299) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (130) to the extremity constituted by nucleotide at position (219) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (130) to the extremity constituted by nucleotide at position (1104) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (130) to the extremity constituted by nucleotide at position (1299) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (220) to the extremity constituted by nucleotide at position (1104) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (220) to the extremity constituted by nucleotide at position (1299) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (1104) to the extremity constituted by nucleotide at position (1299) represented in fig. 5.

The invention also relates to any recombinant nucleic acids containing at least a nucleic acid of the invention inserted in an heterologous nucleic acid.

The invention relates more particularly to recombinant nucleic acid such as defined, in which the nucleotide sequence of the invention is preceded by a promoter (particularly an inducible promoter) under the control of which the transcription of said sequence is liable to be processed and possibly followed by a sequence coding for transcription termination signals.

The invention also relates to the recombinant nucleic acids in which the nucleic acid sequences coding for the polypeptide of the invention and possibly the signal peptide, are recombined with control elements which are heterologous with respect to the ones to which they are normally associated within the bacteria gene and, more particularly, the regulation elements adapted to control their expression in the cellular host which has been chosen for their production.

The invention also relates to recombinant vectors, particularly for cloning and/or expression, comprising a vector sequence, notably of the type plasmid, cosmid or phage, and a recombinant nucleic acid of the invention, in one of the non essential sites for its replication.

Appropriate vectors for expression of the recombinant antigen are the following one:

pEX1 pMTNF MPH  
pEX2 pIGRI  
pEX3  
pUEX1  
pUEX2  
pUEX3

The pEX1, pEX2 and pEX3 vectors are commercially available and can be obtained from Boehringer Mannheim.

The pUEX1, pUEX2 and pUEX3 vectors are also commercially available and can be obtained from Amersham.

According to an advantageous embodiment of the invention, the recombinant vector contains, in one of its non essential sites for its replication, necessary elements to promote the expression of polypeptides according to the invention in a cellular host and possibly a promoter recognized by the polymerase of the cellular host, particularly an inducible promoter and possibly a signal sequence and/or an anchor sequence.

According to another additional embodiment of the invention, the recombinant vector contains the elements enabling the expression by E. coli of a nucleic acid according to the invention inserted in the vector, and particularly the elements enabling the expression of the gene or part thereof of  $\beta$ -galactosidase.

The invention also relates to a cellular host which is transformed by a recombinant vector according to the invention, and comprising the regulation elements enabling the expression of the nucleotide sequence coding for the polypeptide according to the invention in this host.

The invention also relates to a cellular host chosen from among bacteria such as E. coli, transformed by a vector as above defined, and defined hereafter in the examples, or chosen from among eukaryotic organism, such as CHO cells, insect cells, Sf9 cells [Spodoptera frugiperda] infected by the virus Ac NPV (Autographa californica nuclear polyhedrosis virus) containing suitable vectors such as pAc 373 pYM1 or pVC3, BmN [Bombyx mori] infected by the virus BmNPV containing suitable vectors such as pBE520 or p89B310.

The invention relates to an expression product of a nucleic acid expressed by a transformed cellular host according to the invention.

The invention also relates to nucleotidic probes, hybridizing with anyone of the nucleic acids or with their complementary sequences,

and particularly the probes chosen among the following nucleotidic sequences gathered in Table 1, and represented in fig. 9.

TABLE 1

Probes A(i), A(ii), A(iii), A(iv) and A(v)

A(i) CAGCTTGTGACAGGGTTCGTGGC  
 A(ii) GGTTCGTGGCGCCGTACG  
 A(iii) CGTCGCGCGCCTAGTGT CGG  
 A(iv) CGGCGCCGT CGGTGGCACGGCGA  
 A(v) CGTCGGCGCGCCCTAGTGT CGG

Probe B

TCGCCC GCCCTGTACCTG

Probe C

GCGCTGACGCTGGCGATCTATC

Probe D

CCGCTGTTGAAACGT CGGG AAG

Probe E

AAGCCGTCGGATCTGGGTGGCAAC

Probes F(i), F(ii), F(iii) and F(iv)

F(i) ACGGC ACTGGGTGCCACGCCAAC  
 F(ii) ACGCCCAACACCGGGCCCGCCGA  
 F(iii) ACGGGC ACTGGGTGCCACGCCAAC  
 F(iv) ACGCCCCAACACCGGGCCCGCGCCCCA  
 or their complementary nucleotidic sequences.

The hybridization conditions can be the following ones:

- hybridization and wash medium: 3 X SSC, 20% formamide  
 (1 X SSC is 0,15 M NaCl, 0.015 M sodium citrate,

pH 7.0);  
- hybridization temperature (HT) and wash temperature

(WT):

(WT) °C:	HT and WT (°C)
A(i)	50
A(ii)	50
A(iii)	52
A(iv)	60
A(v)	52
B	48
C	50
D	45
E	52
F(i)	55
F(ii)	59
F(iii)	55
F(iv)	59

These probes might enable to differentiate M. tuberculosis from other bacterial strains and in particular from the following mycobacteria species:

- *Mycobacterium marinum*, *Mycobacterium scrofulaceum*, *Mycobacterium gordonaiae*, *Mycobacterium szulgai*, *Mycobacterium intracellulare*, *Mycobacterium xenopi*, *Mycobacterium gastri*, *Mycobacterium nonchromogenicum*, *Mycobacterium terrae* and *Mycobacterium triviale*, and more particularly from M. bovis, *Mycobacterium kansasii*, *Mycobacterium avium*, *Mycobacterium phlei* and *Mycobacterium fortuitum*.

The invention also relates to DNA or RNA primers which can be used for the synthesis of nucleotidic sequences according to the invention by PCR (polymerase chain reaction technique), such as described in US Patents n° 4,683,202 and n° 4,683,195 and European Patent n° 200362.

The invention also relates to any DNA or RNA primer constituted by about 15 to about 25 nucleotides

of a nucleotide sequence coding for a polypeptide according to the invention.

The invention also relates to any DNA or RNA primer constituted by about 15 to about 25 nucleotides liable to hybridize with a nucleotide sequence coding for a polypeptide according to the invention.

The invention also relates to any DNA or RNA primer constituted by about 15 to about 25 nucleotides complementary to a nucleotide sequence coding for a polypeptide according to the invention.

The sequences which can be used as primers are given in Table 2 hereafter (sequences P1 to P6 or their complement) and illustrated in fig. 9 :

TABLE 2

P1	GAGTACCTGCAGGTGCCGTCGCCGTCGATGGGCCG
P2	ATCAACACCCCGGCCTTCGAGTGGTAC
P2 compl.	GTACCACTCGAACGCCGGGGTGTGAT
P3	TGCCAGACTTACAAGTGGGA
P3 compl.	TCCCACTTGTAAGTCTGGCA
P4	TCCTGACCAGCGAGCTGCCG
P4 compl.	CGGCAGCTCGCTGGTCAGGA
P5	CCTGATCGGCCTGGCGATGGGTGACGC
P5 compl.	GCGTCACCCATCGCCAGGCCGATCAGG
P6 compl.	GCGCCCCAGTACTCCCAGCTGTGCGT

compl. = complement

The sequences can be combined in twelve different primer-sets (given in Table 3) which allow enzymatical amplification by the polymerase chain reaction (PCR) technique of any of the nucleotide sequences of the invention, and more particularly the one extending from the extremity constituted by nucleotide at position 1 to the extremity constituted by nucleotide at position 1358, as well as the nucleotide sequence of antigen  $\alpha$  of BCG (17).

The detection of the PCR amplified product can be achieved by a hybridization reaction with an oligonucleotide sequence of at least 10 nucleotides which is located between PCR primers which have been used to amplify the DNA.

The PCR products of the nucleotide sequences of the invention can be distinguished from the  $\alpha$ -antigen gene of BCG or part thereof by hybridization techniques (dot-spot, Southern blotting, etc.) with the probes indicated in Table 3. The sequences of these probes can be found in Table 1 hereabove.

TABLE 3

<u>Primer set</u>	<u>Detection with probe</u>
1. P1 and the complement of P2	B
2. P1 and the complement of P3	B
3. P1 and the complement of P4	B
4. P1 and the complement of P5	B or C
5. P1 and the complement of P6	B, C, D or E
6. P2 and the complement of P5	C
7. P2 and the complement of P6	C, D or E
8. P3 and the complement of P5	C
9. P3 and the complement of P6	C, D or E
10. P4 and the complement of P5	C
11. P4 and the complement of P6	C, D or E
12. P5 and the complement of P6	D or E

It is to be noted that enzymatic amplification can also be achieved with all oligonucleotides with sequences of about 15 consecutive bases of the primers given in Table 2. Primers with elongation at the 5'-end or with a small degree of mismatch may not considerably affect the outcome of the enzymatic amplification if the mismatches do not interfere with the base-pairing at the 3'-end of the primers.

Specific enzymatic amplification of the nucleotide sequences of the invention and not of the BCG gene can be achieved when the probes (given in Table 1) or their complements are used as amplification primers.

When the above mentioned probes of Table 1 are used as primers, the primer sets are constituted by any of the nucleotide sequences (A, B, C, D, E, F) of Table 1 in association with the complement of any other nucleotide sequence, chosen from A, B, C, D, E or F, it being understood that sequence A means any of the sequences A(i), A(ii), A(iii), A(iv), A(v) and sequence F, any of the sequences F(i), F(ii), F(iii) and F(iv).

Advantageous primer sets for enzymatic amplification of the nucleotide sequence of the invention can be one of the following primer sets given in Table 3bis hereafter:

TABLE 3BIS

A(i)	
or A(ii)	
or A(iii)	and the complement of B
or A(iv)	
or A(v)	
A(i)	
or A(ii)	
or A(iii)	and the complement of C
or A(iv)	
or A(v)	
B	and the complement of C
A(i)	
or A(ii)	
or A(iii)	and the complement of F
or A(iv)	
or A(v)	

A(i)  
or A(ii)  
or A(iii) and the complement of D  
or A(iv)  
or A(v)  
A(i)  
or A(ii)  
or A(iii) and the complement of E  
or A(iv)  
or A(v)  
B and the complement of D  
B and the complement of E  
B and the complement of F  
C and the complement of D  
C and the complement of E  
C and the complement of F  
D and the complement of E  
D and the complement of F  
E and the complement of F

A(i), A(ii), A(iii), A(iv), A(v), B, C, D, E and F  
having the nucleotide sequence indicated in Table 1.

In the case of amplification of a nucleotide sequence of the invention with any of the above mentioned primer sets defined in Table 3bis hereabove, the detection of the amplified nucleotide sequence can be achieved by a hybridization reaction with an oligonucleotide sequence of at least 10 nucleotides, said sequence being located between the PCR primers which have been used to amplify the nucleotide sequence. An oligonucleotide sequence located between said two primers can be determined from figure 9 where the primers A, B, C, D, E and F are represented by the boxed sequences respectively named probe region A, probe region B, probe region C, probe region D, probe region E and probe region F.

The invention also relates to a kit for enzymatic amplification of a nucleotide sequence by PCR technique and detection of the amplified nucleotide sequence containing

- one of the PCR primer sets defined in Table 3 and one of the detection probes of the invention, advantageously the probes defined in Table 1, or one of the PCR primer sets defined in Table 3bis, and a detection sequence consisting for instance in an oligonucleotide sequence of at least 10 nucleotides, said sequence being located (fig. 9) between the two PCR primers constituting the primer set which has been used for amplifying said nucleotide sequence.

The invention also relates to a process for preparing a polypeptide according to the invention comprising the following steps:

- the culture in an appropriate medium of a cellular host which has previously been transformed by an appropriate vector containing a nucleic acid according to the invention,
- the recovery of the polypeptide produced by the above said transformed cellular host from the above said culture medium, and
- the purification of the polypeptide produced, eventually by means of immobilized metal ion affinity chromatography (IMAC).

The polypeptides of the invention can be prepared according to the classical techniques in the field of peptide synthesis.

The synthesis can be carried out in homogeneous solution or in solid phase.

For instance, the synthesis technique in homogeneous solution which can be used is the one described by Houbenweyl in the book titled "Methode der organischen Chemie" (Method of organic chemistry)

edited by E. Wunsh, vol. 15-I et II. THIEME, Stuttgart 1974.

The polypeptides of the invention can also be prepared according to the method described by R.D. MERRIFIELD in the article titled "Solid phase peptide synthesis" (J.P. Ham.Socks. , 45, 2149-2154).

The invention also relates to a process for preparing the nucleic acids according to the invention.

A suitable method for chemically preparing the single-stranded nucleic acids (containing at most 100 nucleotides of the invention) comprises the following steps :

- DNA synthesis using the automatic  $\beta$ -cyanoethyl phosphoramidite method described in Bioorganic Chemistry 4; 274-325, 1986.

In the case of single-stranded DNA, the material which is obtained at the end of the DNA synthesis can be used as such.

A suitable method for chemically preparing the double-stranded nucleic acids (containing at most 100 bp of the invention) comprises the following steps:

- DNA synthesis of one sense oligonucleotide using the automatic  $\beta$ -cyanoethyl phosphoramidite method described in Bioorganic Chemistry 4; 274-325, 1986, and DNA synthesis of one anti-sense oligonucleotide using said above-mentioned automatic  $\beta$ -cyanoethyl phosphoramidite method,

- combining the sense and anti-sense oligonucleotides by hybridization in order to form a DNA duplex,

- cloning the DNA duplex obtained into a suitable plasmid vector and recovery of the DNA according to classical methods, such as restriction enzyme digestion and agarose gel electrophoresis.

A method for the chemical preparation of nucleic acids of length greater than 100 nucleotides - or bp,

in the case of double-stranded nucleic acids - comprises the following steps :

- assembling of chemically synthesized oligonucleotides, provided at their ends with different restriction sites, the sequences of which are compatible with the succession of amino acids in the natural peptide, according to the principle described in Proc. Nat. Acad. Sci. USA 80; 7461-7465, 1983,

- cloning the DNA thereby obtained into a suitable plasmid vector and recovery of the desired nucleic acid according to classical methods, such as restriction enzyme digestion and agarose gel electrophoresis.

The invention also relates to antibodies themselves formed against the polypeptides according to the invention.

It goes without saying that this production is not limited to polyclonal antibodies.

It also relates to any monoclonal antibody produced by any hybridoma liable to be formed according to classical methods from splenic cells of an animal, particularly of a mouse or rat, immunized against the purified polypeptide of the invention on the one hand, and of cells of a myeloma cell line on the other hand, and to be selected by its ability to produce the monoclonal antibodies recognizing the polypeptide which has been initially used for the immunization of the animals.

The invention also relates to any antibody of the invention labeled by an appropriate label of the enzymatic, fluorescent or radioactive type.

The peptides which are advantageously used to produce antibodies, particularly monoclonal antibodies, are the following ones gathered in Table 4:

TABLE 4a (see fig. 4a and 4b)

Amino acid position (NH <sub>2</sub> -terminal)	Amino acid position (COOH-terminal)
12	QVPSPSMGRDIKVQFQSGGA 31
36	LYLLDGLRAQDDFSGWDINT 55
77	SFYSDWYQPACRKAGCQTYK 96
101	LTSELPGWLQANRHVKPTGS 120
175	KASDMWGPKEPDPAWQRNDPL 194
211	CGNGKPSDLGGNNLPAKFLE 230
275	KPDLQRHWVPRPTPGPPQGA 294

TABLE 4b (see fig. 5)

Amino acid position (NH <sub>2</sub> -terminal)	Amino acid position (COOH-terminal)
77	SFYSDWYQPACGKAGCQTYK 96
276	PDLQRALGATPNTGPAPQGA 295

The amino acid sequences are given in the 1-letter code.

Variations of the peptides listed in Table 4 are also possible depending on their intended use. For example, if the peptides are to be used to raise antisera, the peptides may be synthesized with an extra cysteine residue added. This extra cysteine residue is preferably added to the amino terminus and facilitates the coupling of the peptide to a carrier protein which is necessary to render the small peptide immunogenic. If the peptide is to be labeled for use in radioimmune assays, it may be advantageous to synthesize the protein with a tyrosine attached to either the amino or carboxyl terminus to facilitate iodination. These

peptides possess therefore the primary sequence of the peptides listed in Table 4 but with additional amino acids which do not appear in the primary sequence of the protein and whose sole function is to confer the desired chemical properties to the peptides.

The invention also relates to a process for detecting in vitro antibodies related to tuberculosis in a human biological sample liable to contain them, this process comprising

- contacting the biological sample with a polypeptide or a peptide according to the invention under conditions enabling an in vitro immunological reaction between said polypeptide and the antibodies which are possibly present in the biological sample and
- the in vitro detection of the antigen/antibody complex which may be formed.

Preferably, the biological medium is constituted by a human serum.

The detection can be carried out according to any classical process.

By way of example a preferred method brings into play an immunoenzymatic process according to ELISA technique or immunofluorescent or radioimmunological (RIA) or the equivalent ones.

Thus the invention also relates to any polypeptide according to the invention labeled by an appropriate label of the enzymatic, fluorescent, radioactive... type.

Such a method for detecting in vitro antibodies related to tuberculosis comprises for instance the following steps:

- deposit of determined amounts of a polypeptidic composition according to the invention in the wells of a titration microplate,
- introduction into said wells of increasing dilutions of the serum to be diagnosed,

- incubation of the microplate,
- repeated rinsing of the microplate,
- introduction into the wells of the microplate of labeled antibodies against the blood immunoglobulins,
- the labeling of these antibodies being carried out by means of an enzyme which is selected from among the ones which are able to hydrolyze a substrate by modifying the absorption of the radiation of this latter at least at a given wave length,
- detection by comparing with a control standard of the amount of hydrolyzed substrate.

The invention also relates to a process for detecting and identifying in vitro antigens of M. tuberculosis in a human biological sample liable to contain them, this process comprising:

- contacting the biological sample with an appropriate antibody of the invention under conditions enabling an in vitro immunological reaction between said antibody and the antigens of M. tuberculosis which are possibly present in the biological sample and the in vitro detection of the antigen/antibody complex which may be formed.

Preferably, the biological medium is constituted by sputum, pleural effusion liquid, broncho-alveolar washing liquid, urine, biopsy or autopsy material.

Appropriate antibodies are advantageously monoclonal antibodies directed against the peptides which have been mentioned in Table 4.

The invention also relates to an additional method for the in vitro diagnostic of tuberculosis in a patient liable to be infected by *Mycobacterium tuberculosis* comprising the following steps:

- the possible previous amplification of the amount of the nucleotide sequences according to the invention, liable to be contained in a biological sample taken

from said patient by means of a DNA primer set as above defined,

- contacting the above mentioned biological sample with a nucleotide probe of the invention, under conditions enabling the production of an hybridization complex formed between said probe and said nucleotide sequence,
- detecting the above said hybridization complex which has possibly been formed.

To carry out the in vitro diagnostic method for tuberculosis in a patient liable to be infected by *Mycobacterium tuberculosis* as above defined, the following necessary or kit can be used, said necessary or kit comprising:

- a determined amount of a nucleotide probe of the invention,
- advantageously the appropriate medium for creating an hybridization reaction between the sequence to be detected and the above mentioned probe;
- advantageously, reagents enabling the detection of the hybridization complex which has been formed between the nucleotide sequence and the probe during the hybridization reaction.

The invention also relates to an additional method for the in vitro diagnostic of tuberculosis in a patient liable to be infected by *Mycobacterium tuberculosis* comprising :

- contacting a biological sample taken from a patient with a polypeptide or a peptide of the invention, under conditions enabling an in vitro immunological reaction between said polypeptide or peptide and the antibodies which are possibly present in the biological sample and
- the in vitro detection of the antigen/antibody complex which has possibly been formed.

To carry out the in vitro diagnostic method for tuberculosis in a patient liable to be infected by

Mycobacterium tuberculosis, the following necessary or kit can be used, said necessary or kit comprising:

- a polypeptide or a peptide according to the invention,
- reagents for making a medium appropriate for the immunological reaction to occur,
- reagents enabling to detect the antigen/antibody complex which has been produced by the immunological reaction, said reagents possibly having a label, or being liable to be recognized by a labeled reagent, more particularly in the case where the above mentioned polypeptide or peptide is not labeled.

The invention also relates to an additional method for the in vitro diagnostic of tuberculosis in a patient liable to be infected by M. tuberculosis, comprising the following steps:

- contacting the biological sample with an appropriate antibody of the invention under conditions enabling an in vitro immunological reaction between said antibody and the antigens of M. tuberculosis which are possibly present in the biological sample and - the in vitro detection of the antigen/antibody complex which may be formed.

Appropriate antibodies are advantageously monoclonal antibodies directed against the peptides which have been mentioned in Table 4.

To carry out the in vitro diagnostic method for tuberculosis in a patient liable to be infected by Mycobacterium tuberculosis, the following necessary or kit can be used, said necessary or kit comprising:

- an antibody of the invention,
- reagents for making a medium appropriate for the immunological reaction to occur,
- reagents enabling to detect the antigen/antibody complexes which have been produced by the immunological reaction, said reagent possibly having a label or being

liable to be recognized by a label reagent, more particularly in the case where the above mentioned antibody is not labeled.

An advantageous kit for the diagnostic in vitro of tuberculosis comprises:

- at least a suitable solid phase system, e.g. a microtiter-plate for deposition thereon of the biological sample to be diagnosed in vitro,
- a preparation containing one of the monoclonal antibodies of the invention,
- a specific detection system for said monoclonal antibody,
- appropriate buffer solutions for carrying out the immunological reaction between a test sample and said monoclonal antibody on the one hand, and the bonded monoclonal antibodies and the detection system on the other hand.

The invention also relates to a kit, as described above, also containing a preparation of one of the polypeptides or peptides of the invention, said antigen of the invention being either a standard (for quantitative determination of the antigen of M. tuberculosis which is sought) or a competitor, with respect to the antigen which is sought, for the kit to be used in a competition dosage process.

The invention also relates to an immunogenic composition comprising a polypeptide or a peptide according to the invention, in association with a pharmaceutically acceptable vehicle.

The invention also relates to a vaccine composition comprising among other immunogenic principles anyone of the polypeptides or peptides of the invention or the expression product of the invention, possibly coupled to a natural protein or to a synthetic polypeptide having a sufficient molecular weight so that the conjugate is able to induce in vivo

the production of antibodies neutralizing *Mycobacterium tuberculosis*, or induce *in vivo* a cellular immune response by activating *M. tuberculosis* antigen-responsive T cells.

The peptides of the invention which are advantageously used as immunogenic principle have one of the following sequences:

TABLE 4a (see fig. 4a and 4b)

Amino acid position (NH <sub>2</sub> -terminal)	Amino acid sequence	Amino acid position (COOH-terminal)
12	QVPSPSMGRDIKVQFQSGGA	31
36	LYLLDGLRAQDDFSGWDINT	55
77	SFYSDWYQPACRKAGCQTYK	96
101	LTSELPGWLQANRHVKPTGS	120
175	KASDMWGPKEDEPAWQRNDPL	194
211	CGNGKPSDLGGNNLPAKFLE	230
275	KPDLQRHWVPRPTPGPPQGA	294

TABLE 4b (see fig. 5)

Amino acid position (NH <sub>2</sub> -terminal)	Amino acid sequence	Amino acid position (COOH-terminal)
77	SFYSDWYQPACGKAGCQTYK	96
276	PDLQRALGATPNTGPAPQGA	299

The amino acid sequences are given in the 1-letter code.

Other characteristics and advantages of the invention will appear in the following examples and the figures illustrating the invention.

Figures 1(A) and 1(B) correspond to the identification of six purified *λgt11 M. tuberculosis* recombinant clones. Figure 1(A) corresponds to the

EcoRI restriction analysis of clone 15, clone 16, clone 17, clone 19, clone 24 and EcoRI-HindIII digested lambda DNA-molecular weight marker lane (in kilobase pairs) (M) (Boehringer).

Figure 1(B) corresponds to the immunoblotting analysis of crude lysates of *E. coli* lysogenized with clone 15, clone 16, clone 17, clone 19, clone 23 and clone 24.

Arrow (←) indicates fusion protein produced by recombinant  $\lambda$ gt11-*M. tuberculosis* clones. Expression and immunoblotting were as described above. Molecular weight (indicated in kDa) were estimated by comparison with molecular weight marker (High molecular weight-SDS calibration kit, Pharmacia).

Figure 2 corresponds to the restriction map of the DNA inserts in the  $\lambda$ gt11 *M. tuberculosis* recombinant clones 17 and 24 identified with polyclonal anti-32-kDa (BCG) antiserum as above defined and of clones By1, By2 and By5 selected by hybridization with a 120 bp EcoRI-Kpn I restriction fragment of clone 17.

DNA was isolated from  $\lambda$ gt11 phage stocks by using the Lambda Sorb Phage Immunoadsorbent, as described by the manufacturer (Promega). Restriction sites were located as described above. Some restriction sites (\*) were deduced from a computer analysis of the nucleotide sequence.

The short vertical bars (|—|) represent linker derived EcoRI sites surrounding the DNA inserts of recombinant clones. The lower part represents a magnification of the DNA region containing the antigen of molecular weight of 32-kDa, that has been sequenced. Arrows indicate strategies and direction of dideoxy-sequencing. (→) fragment subcloned in Bluescribe M13; (←) fragment subcloned in mp10 and mp11 M13 vectors; (→) sequence determined with the use of a synthetic oligonucleotide.

Figures 3a and 3b correspond to the nucleotide and amino acid sequences of the general formula of the antigens of the invention.

Figures 4a and 4b correspond to the nucleotide and amino acid sequences of one of the antigens of the invention.

Two groups of sequences resembling the *E. coli* consensus promoter sequences are boxed and the homology to the consensus is indicated by italic bold letters. Roman bold letters represent a putative Shine-Dalgarno motif.

The NH<sub>2</sub>-terminal amino acid sequence of the mature protein which is underlined with a double line happens to be very homologous - 29/32 amino acids - with the one of MPB 59 antigen (34). Five additional ATG codons, upstream of the ATG at position 273 are shown (dotted underlined). Vertical arrows (↓) indicate the presumed NH<sub>2</sub> end of clone 17 and clone 24. The option taken here arbitrarily represents the 59 amino acid signal peptide corresponding to ATG<sub>183</sub>.

Figure 5 corresponds to the nucleotide and amino acid sequences of the antigen of 32-kDa of the invention.

The NH<sub>2</sub>-terminal amino acid sequence of the mature protein which is underlined with a double line happens to be very homologous - 29/32 amino acids - with the one of MPB 59 antigen (34). Vertical arrows (↓) indicate the presumed NH<sub>2</sub> end of clone 17 and clone 24.

Figure 6 is the hydropathy pattern of the antigen of the invention of a molecular weight of 32-kDa and of the antigen  $\alpha$  of BCG (17).

Figure 7 represents the homology between the amino acid sequences of the antigen of 32-kDa of the invention and of antigen  $\alpha$  of BCG (revised version).

Identical amino acids; (:) evolutionarily conserved replacement of an amino acid (:), and absence

of homology () are indicated. Underlined sequence (=) represents the signal peptide, the option taken here arbitrarily representing the 43-amino acid signal peptide corresponding to ATG<sub>11</sub>. Dashes in the sequences indicate breaks necessary for obtaining the optimal alignment:

Figure 8 illustrates the fact that the protein of 32-kDa of the invention is selectively recognized by human tuberculous sera.

Figure 8 represents the immunoblotting with human tuberculous sera, and anti- $\beta$ -galactosidase monoclonal antibody. Lanes 1 to 6: E. coli lysate expressing fusion protein (140 kDa); lanes 7 to 12: unfused  $\beta$ -galactosidase (114 kDa). The DNA insert of clone 17 (2.7 kb) was subcloned into pUEX<sub>2</sub> and expression of fusion protein was induced as described by Bresson and Stanley (4). Lanes 1 and 7 were probed with the anti- $\beta$ -galactosidase monoclonal antibody: lanes 4, 5, 6 and 10, 11, 12 with 3 different human tuberculous sera highly responding towards purified protein of the invention of 32-kDa; lanes 2 and 3 and 8 and 9 were probed with 2 different low responding sera.

Figure 9 represents the nucleic acid sequence alignment of the 32-kDa protein gene of *M. tuberculosis* of the invention (upper line), corresponding to the sequence in fig. 5, of the gene of fig. 4a and 4b of the invention (middle line), and of the gene for antigen  $\alpha$  of BCG (lower line).

Dashes in the sequence indicate breaks necessary for obtaining optimal alignment of the nucleic acid sequence.

The primer regions for enzymatical amplification are boxed (P1 to P6).

The specific probe regions are boxed and respectively defined by probe region A, probe region B,

probe region C, probe region D, probe region E and probe region F.

It is to be noted that the numbering of nucleotides is different from the numbering of figures 3a and figure 3b, and of figure 5, because nucleotide at position 1 (on figure 9) corresponds to nucleotide 234 on Figure 3a, and corresponds to nucleotide 91 on figure 5.

Figure 10a corresponds to the restriction and genetic map of the pIGRI plasmid used in Example IV for the expression of the  $P_{32}$  antigen of the invention in *E. coli*.

On this figure, underlined restriction sites are unique.

Figure 10b corresponds to the pIGRI nucleic acid sequence.

On this figure, the origin of nucleotide stretches used to construct plasmid pIGRI are specified hereafter.

Position

3422-206 :	lambda PL containing EcoRI blunt-MboII blunt fragment of pPL( $\lambda$ ) (Pharmacia)
207-384 :	synthetic DNA sequence
228-230 :	initiation codon ATG of first cistron
234-305 :	DNA encoding amino acids 2 to 25 of mature mouse TNF
306-308 :	stop codon (TAA) first cistron
311-312 :	initiation codon (ATG) second cistron
385-890 :	rrnBT <sub>1</sub> T <sub>2</sub> containing HindIII-SspI fragment from pKK223 (Pharmacia)
891-3421 :	DraI-EcoRI blunt fragment of pAT <sub>153</sub> (Bioexcellence) containing the

tetracycline resistance gene and the origin of replication.

Table 5 hereafter corresponds to the complete restriction site analysis of pIGRI.

---

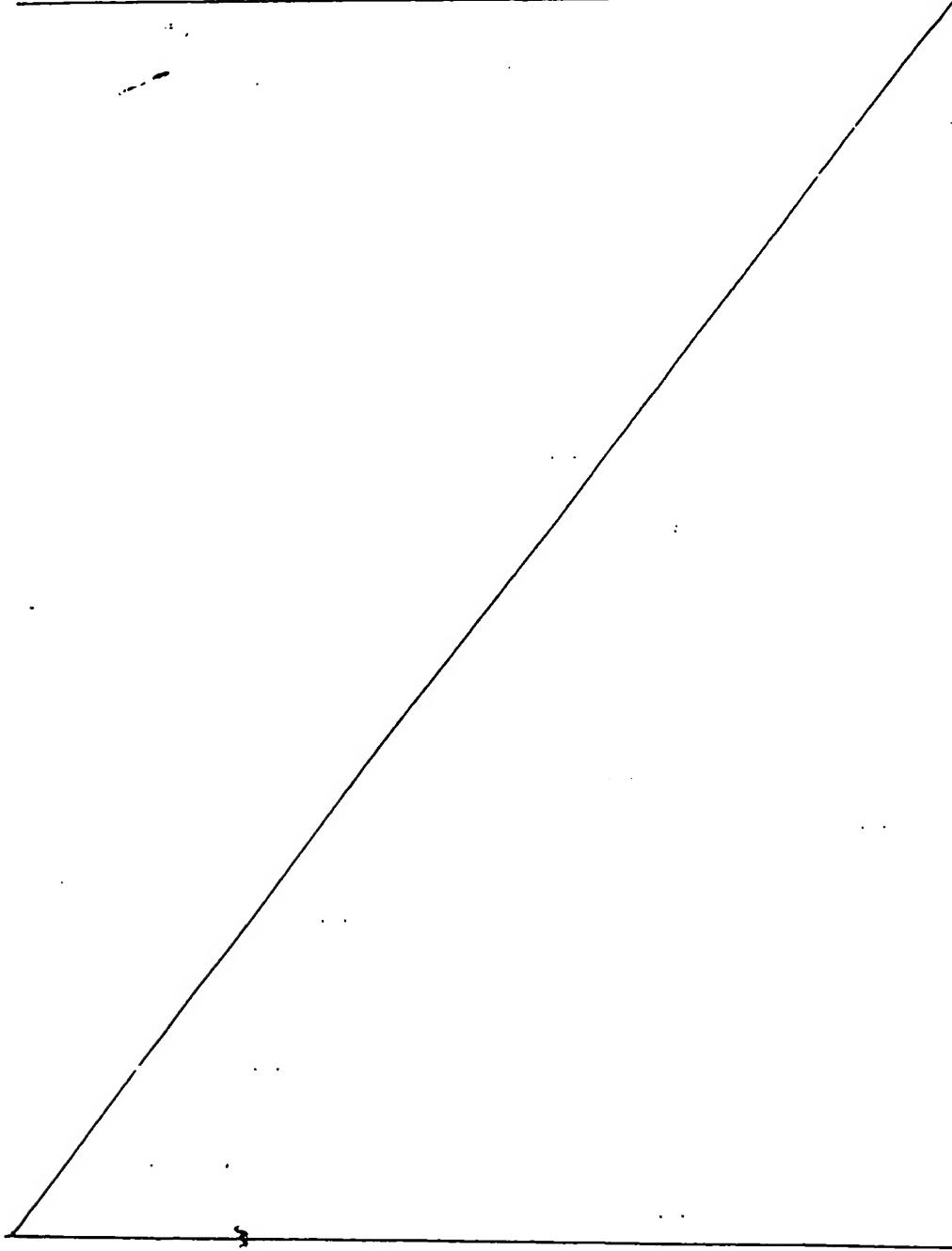


Table 5

\*\*\*\*\*  
\* RESTRICTION-SITE ANALYSIS \*  
\*\*\*\*\*

Name of the plasmid : PIGRI

Total number of bases is: 3423.  
Analysis done on the complete sequence.

List of cuts by enzyme.

Acc I	:	370	2765		
AcY I	:	735	2211	2982	3003
AfI III	:	1645			
Aha III	:	222			
Alu I	:	386	1088	1345	1481
Alw NI	:	1236			
Apa LI	:	1331			
Asp 718I	:	208			
Asu I	:	329	494	623	713
Ava I	:	3244			
Ava II	:	1990			
Bal I	:	329	494	1935	1977
Bam HI	:	1973			
Bbe I	:	3040			
Bbv I	:	2214	2871	2985	3006
Bbv I*	:	389	1316	1735	1753
Bbv I*	:	1017	1223	1226	1973
Bbv II	:	1822	2685		
Bgl I	:	2253	2487		
Bin I	:	15	903	1001	1087
Bin I*	:	902	999	2313	3035

Table 5 (Con't.)

		855	925	2926				
Bsp	HI	382	2361					
Bsp	MI	213	475	585	753	1486	1499	1620
Bst	NI	4	683	716	1268	1933	2159	2883
Cau	II	2132	2486	2646	3005	3014	3255	3247
Cfr	10I	1971	2476	2884	3016	3120		
Cfr	I	3393						
Cla	I							
Cvi	JI	190	263	270	380	386	391	421
		791	1088	1117	1160	1171	1236	1315
		1605	1623	1634	1707	1726	1926	1931
		2157	2162	2300	2310	2329	2370	2427
		2544	2588	2732	2748	2804	2822	2886
		3087	3122	3245	3269	3388	3403	2894
Cvi	QI	209	3253					
Dde	I	133	336	343	518	608	664	625
Dpn	I	9	236	897	909	987	995	1006
Dra	II	1935	1977	2592	2951	3042	3069	1081
Dra	III	293						
Dsa	I	309	1968	2887				
Eco	3II	562						
Eco	47III	341	1773	2642	2923	3185		
Eco	57I	214						
Eco	57I*	1103						
Eco	78I	2212	2869	2983	3004			
Eco	NI	196	2792					
Eco	RII	211	473	583	751	1484	1497	1618
Eco	RV	3232						
Fnu	4HI	378	479	1031	1237	1240	1305	1448
Fnu		1855	1858	1987	2001	2008	2011	2130
		2479	2644	2695	2802	2836	2839	3117
Fnu	DII	489	1021	1602	1784	1881	2003	2029
Fok	I	2440	2445	2472	2601	2716	3072	2174
Fok	I*	415	799	3317				
Gsu	I	763	2370	2415	3269			
Gsu	I*	339	2035	2589				

Table 5 (Con't)

Hae I	:	775	791	1171	1623	1634	1973	2370	2427	2499
Hae II	:	343	541	1405	1775	2214	2644	2871	2925	2985
Hae III	:	625	714	775	791	1171	1605	1623	1634	1973
Hgq I	:	2427	2478	2499	2588	2822	2886	2894	3018	3122
Hgq I*	:	158	181	743	2035	2185	2776			3245
Hgq II	:	955	1533	2429	2461	3015				
Hgq AI	:	139	1335	1954	2245	2832	3143			
Hgq CI	:	208	2126	2210	2649	2867	2981	3002	3296	3339
Hgq III	:	2934	2948							
Bha I	:	342	489	540	1021	1130	1304	1404	1471	1774
Bpa II	:	2000	2062	2213	2472	2603	2643	2718	2870	2924
Bpa III	:	3158	3186	3318						2984
Bin P11	:	340	487	538	1019	1128	1302	1402	1469	1739
Bind II	:	1998	2060	2211	2470	2601	2641	2716	2868	2922
Bind III	:	3156	3184	3316						2982
Hinf I	:	367	1275	1671	1746	1891	2112	2410	2564	2784
Hph I	:	3	682	716	1077	1267	1293	1440	1932	2133
Hph II	:	2487	2647	2723	2883	3006	3015	3030	3247	3256
Hph I*	:	94	138	181	663	914	1900	2121	2975	3020
Kpn I	:	6	212							3302
Mae I	:	364	899	1152	1928	3187				
Mae II	:	274	698	944	1847	1871	2460	2516		
Mae III	:	169	255	304	313	1109	1225	1288	2267	2534
Mbo I	:	7	234	895	907	985	993	1004	1079	1955
Mbo II	:	2318	2590	2949	3040	3067				
Mbo II*	:	207	422	917	1779	1827	2419	2690		
Mme I*	:	988	2944							
Mnl I	:	1252	1436	3112	3199					
Mnl I*	:	1218	1542	1948	2446	2630				
Mse I	:	208	289	337	711	1467	1750	2116	2143	2181
Mst I	:	2811	3030	3234	3294					2242
Nae I	:	179	186	221	433	764	941	3361	3383	3420
		1963	2061	3157						
		2134	2488	2648	3016					

Table 5 (Con't)

		2211	2868	2982	3003	
Nar I	:	309				
Nco I	:	3186	230	313	512	567
Nhe I	:	166	2241	2369	2486	2672
Nla III	:	2226	210	330	496	1578
Nla IV	:	2530	2445	2651	2869	2983
Nru I	:					
Nsp BII	:	1062	1307	2278		
Nsp HI	:	1649	2857			
PF1 MI	:	293	2052	2101		
Ple I	:	375	1754			
Ple I*	:	1269	2778			
Ppu MI	:	1935	1977			
Pss I	:	1938	1980	2895		
Pst I	:	379				
Rsa I	:	210	3254			
Sal I	:	369	2764			
Scr FI	:	4	213	475	585	683
Sdu I	:	1933	1975	2159	2358	2883
Sec I	:	139	1335	1954	2245	2832
Sfa NI	:	3	309	1485	1968	2046
Sfa NI*	:	597	765	2392	2767	3178
Sph I	:	1548	1985	2380	3001	3013
Sso I	:	2857				
Sty I	:	2	211	473	583	681
Taq I	:	1931	1973	2157	2356	2881
Taq IIB	:	309	2046			
Taq IIB*	:	252	370	613	1547	2149
TthIIII*	:	1749	2751			
Xba I	:	38	1054			
Xho II	:	633	1022	1061		
Xma III	:	363				
Xmn I	:	7	895	907	993	1004
		2476				3040
						414
						Total number of cuts is : 705.

Table 5 (Con't)

Sorted list of enzymes by n° of cuts.

Cvi JI	:	61	Sdu I	:	8	Tth1111*	:	1
Fnu 4BI	:	31	Cau I	:	8	Nsp DII	:	3
Bfa I	:	25	Bbv I	:	8	Fok I	:	3
Bin PII	:	25	Mbo I	:	7	Pfl MI	:	3
Rae III	:	21	Ava I	:	7	Rind II	:	3
Mla IV	:	21	Mae I	:	7	Dga I	:	3
Nla III	:	21	Sfa NI	:	6	Bsp HI	:	3
Hpa II	:	20	Xho II	:	6	Pss I	:	3
Scr FI	:	18	Bgi AI	:	6	Hst I	:	3
Sso II	:	18	Sfa NI*	:	6	Rgi JII	:	2
Fnu DII	:	17	Bbv I*	:	6	Ple I	:	2
Mbo I	:	16	Cfr 10I	:	6	Hbo III*	:	2
Dpn I	:	16	Bga I	:	6	Cvi QI	:	2
Mnl I*	:	15	Acy I	:	5	Acc I	:	2
Asu I	:	12	Bin I	:	5	Bgl I	:	2
Bae II	:	11	Cfr I	:	5	Ple I*	:	2
Mae III	:	11	Hga I*	:	5	Gsu I	:	2
Rph I	:	10	Hae I	:	5	PPu MI	:	2
Bst NI	:	10	Eco 47III	:	5	Tth1111*	:	2
Eco RII	:	10	Mnl I	:	5	Rind III	:	2
Sec I	:	10	Hme I*	:	4	Nsp BI	:	2
Dde I	:	9	Eco 78I	:	4	Rsa I	:	2
Rinf I	:	9	Nae I	:	4	Sal I	:	2
Hae I	:	9	Bbs I	:	4	Bbv II	:	2
Alu I	:	9	Bin I*	:	4	Bsp HI	:	2
Bgi CI	:	9	Nar I	:	4	Sty I	:	2
Hse I	:	9	Fok I*	:	4	Eco NI	:	2
Taq I	:	9	Dra I	:	3	Xmn I	:	2

## List of non cutting selected enzymes:

At II	,	Afl II	,	Apa I	,	Asu II	,	Avr II	,	Bbv II*	,	Bcl I
Bql II	,	Bsp MI*	,	Bsp MII	,	Bss III	,	Bst EII	,	Bst XI	,	Eco 31I*
Eco RI	,	Esp I	,	Hpa I	,	Mlu I	,	Mme I	,	Nde I	,	Not I
Nsi I	,	Pma CI	,	Pvu I	,	Pvu II	,	Rsr II	,	Sac I	,	Sac II
Sau I	,	Sca I	,	Sci I	,	Sfi I	,	Sma I	,	Sna BI	,	Spe I
Spl I	,	Ssp I	,	Stu I	,	Taq IIA	,	Taq IIA*	,	Tth 1111	,	Vsp I
Xca I	,	Xho I	,	Xma I								Total number of selected enzymes which do not cut: 45

Figure 11a corresponds to the restriction and genetic map of the pmTNF MPH plasmid used in Example V for the expression of the  $P_{32}$  antigen of the invention in E. coli.

Figure 11b corresponds to the pmTNF-MPH nucleic acid sequence.

On this figure, the origin of nucleotide stretches used to construct plasmid pmTNF-MPH is specified hereafter.

Position

1-208 : lambda PL containing EcoRI blunt-MboII blunt fragment of pPL( $\lambda$ ) (Pharmacia)  
209-436 : synthetic DNA fragment  
230-232 : initiation codon (ATG) of mTNF fusion protein  
236-307 : sequence encoding AA 2 to 25 of mature mouse TNF  
308-384 : multiple cloning site containing His<sub>6</sub> encoding sequence at position 315-332  
385-436 : HindIII fragment containing E. coli trp terminator  
437-943 : rrnBT<sub>1</sub>T<sub>2</sub> containing HindIII-SspI fragment from pKK223 (Pharmacia)  
944-3474 : DraI-EcoRI blunt fragment of pAT<sub>153</sub> (Bioexcellence) containing the tetracycline resistance gene and the origin of replication.

Table 6 hereafter corresponds to the complete restriction site analysis of pmTNF-MPH.

Table 6

\*\*\*\*\*  
\* RESTRICTION-SITE ANALYSIS \*  
\*\*\*\*\*

Done on DNA sequence PMTNFMPH.

Total number of bases is: 3474.  
Analysis done on the complete sequence.

List of cuts by enzyme.

Acc I	:	371	2818								
AcY I	:	788	2264	2921	3035	3056					
AfI II	:	387									
AfI III	:	1698									
Aha III	:	224									
Alu I	:	386	439	1141	1398	1534	1760	2382	2785	3441	3456
Alw NI	:	1289									
Apa I	:	345									
Apa LI	:	1384									
Asp 718I	:	210									
Asu I	:	341	342	547	676	766	1988	2030	2209	2333	2582
											267
Ava I	:	2945	3297								
Ava II	:	338	2043								
Bal I	:	547	1988	2030	2333	2582	2670				
Bam HI	:	334	3093								
Bbe I	:	2267	2924	3038	3059						
Bbv I	:	1369	1788	1806	1919	1922	2866	3255			
Bbv I*	:	1070	1276	1279	2026	2050	2683				

SUBSTITUTE SHEET

Table 6 (Con't)

Bbv	II	:	1875	2738				
Bgl	I	:	2306	2540				
Bin	I	:	117	342	956	1054	1140	3101
Bin	I*	:	329	955	1052	2366	3088	
Bsp	HI	:	908	978	2979			
Bsp	MI	:	2414					
Bsp	MII	:	354					
Bst	NI	:	215	528	638	806	1539	1552
Cau	II	:	6	339	340	736	769	1321
Cfr	10I	:	374	2185	2539	2699	3058	3067
Cfr	I	:	2024	2529	2937	3069	3173	3308
Cla	I	:	3446					
Cvi	JII	:	192	265	272	343	350	361
			660	678	767	828	844	1141
			1393	1398	1534	1632	1658	1676
			2026	2063	2145	2189	2210	2215
			2488	2518	2531	2552	2597	2641
			2947	2985	2999	3071	3140	3175
Cvi	QI	:	211	3306				
Dde	I	:	135	571	661	717	1015	1424
Dpn	I	:	11	238	336	950	962	1040
Dra	II	:	2342	2373	2645	3004	3095	3122
Dra	III	:	1988	2030	2945			
Dsa	I	:	295	331				
Eco	3II	:	345	2021				
Eco	47III	:	615					
Eco	57I	:	1826	2695	2976	3238		
Eco	57I*	:	216					
Eco	78I	:	1156					
Eco	NI	:	2265	2922	3036	3057		
Eco	RI	:	198	2845				
Eco	RII	:	309					
Eco	RV	:	213	526	636	804	1537	1550
			3285					

Table 6 (con't)

Fnu 4HI	:	401	417	532	1084	1290	1293	1358	1501	1656	1774	177
		1795	1908	1911	2040	2054	2061	2064	2183	2262	2307	236
		2447	2532	2697	2748	2855	2889	2892	3170	3173	3244	
Fnu DII	:	542	1074	1655	1837	1934	2056	2082	2227	2237	2366	243
Fok I	:	2493	2498	2525	2654	2769	3125					
Fok I*	:	468	852	3370	2468	3322						
Gsu I	:	2088										
Gsu I*	:	2642										
Hae I	:	361	828	844	1224	1676	1687	2026	2423	2480	2552	
Hae II	:	594	1458	1828	2267	2697	2924	2978	3038	3059	3240	
Hae III	:	343	361	678	767	828	844	1224	1658	1676	1687	202
		2210	2423	2480	2531	2552	2641	2875	2939	2947	3071	317
Hga I	:	3298										
Hga I*	:	160	183	796	2088	2238	2829					
Hgi AI	:	1008	1586	2482	2514	3068						
Hgi CI	:	141	1388	2007	2298	2885	3196					
Hgi JII	:	210	2179	2263	2702	2920	3034	3055	3349	3392		
Hha I	:	345	2987	3001								
		542	593	1074	1183	1357	1457	1524	1794	1827	2017	205
		2115	2266	2525	2656	2696	2771	2923	2977	3037	3058	321
Hin PII	:	3239	3371									
		540	591	1072	1181	1355	1455	1522	1792	1825	2015	205
		2113	2264	2523	2654	2694	2769	2921	2975	3035	3056	320
Bind II	:	3237	3369									
Bind III	:	109	372	2819								
Bind I	:	384	437	3439								
		368	1328	1724	1799	1944	2165	2463	2617	2837		

SUBSTITUTE SHEET

Table 6 (Con't)

Hpa II	:	5	339	355	375	735	769	1130	1320	1346	1493	198
Hph I*	:	8	305	311	317							
Hph II	:	214										
Kpn I	:											
Mae I	:	365	952	1205	1981	3240						
Mae II	:	276	330	751	997	1900	1924	2513	2569			
Mae III	:	171	257	1162	1278	1341	2320	2587	3255	3343		
Mbo I	:	9	236	334	948	960	1038	1046	1057	1132	2008	232
Mbo II	:	2340	2371	2643	3002	3093	3120					
Mbo II*	:	209	475	970	1832	1880	2472	2743				
Mme I*	:	1041	2997									
Mnl I	:	1305	1489	3165	3252							
Mnl I*	:	372	1271	1595	2001	2499	2683					
Nla I	:	210	291	350	764	1520	1803	2169	2196	2234	2295	259
Nse I	:	2864	3083	3287	3347							
Mst I	:	181	188	223	388	486	817	994	3414	3436		
Nae I	:	2016	2114	3210								
Nar I	:	2187	2541	2701	3069							
Nco I	:	2264	2921	3035	3056							
Nhe I	:	345										
Nla III	:	3239	168	232	349	382	565	620	912	982	1702	1881
												201
Nla IV	:	2222	2279	2294	2422	2539	2725	2764	2910	2983	3121	346
Nru I	:	212	336	343	549	1631	1670	1989	2032	2146	2181	221
Nsp BII	:	2265	2583	2704	2922	2946	3036	3057	3095	3141	3351	339
Nsp HI	:	2498	412	1115	1360	2331						
		382	1702	2910								

79 A

Table 6 (Con't)

		295	2105	2154
Pfl MI	:	376	1807	
Ple I	:	1322	2831	
Ple I*	:	331		
Pma CI	:	1988	2030	
Ppu MI	:	1991	2033	2948
Pss I	:	212	3307	
Rsa I	:	370	2817	
Sal I	:	6	215	339
Scr FI	:	1552	1673	1986
Sdu I	:	141	345	1388
Sec I	:	5	338	345
Sfa NI	:	650	818	2445
Sfa NI*	:	420	1601	2038
Sma I	:	340		
Sph I	:	382	2910	
Sso II	:	4	213	337
Stu I	:	1550	1671	1984
Sty I	:	361		
Taq I	:	345	2099	
Taq IIB	:	254	371	
Taq IIB*	:	1802	666	1600
Tth1111	:	2804		
Tth1111*	:	40	1107	
Xba I	:	686	1075	1114
Xho II	:	364		
Xma I	:	9	334	948
Xma III	:	338		
Xmn I	:	2529		
		467		

Total number of cuts is : 743.

SUBSTITUTE SHEET

Table 6 (con't)

## List of non cutting selected enzymes.

Aat II	Asu II	Avr II	Bbv II*	Bcl I	Bgl II	Bsp MI*
Bss HII	Bst EII	Bst XI	Eco 31I*	Esp I	Hpa I	Mlu I
Mme I	Nde I	Not I	Nsi I	Pst I	Pvu I	Pvu II
Rsr II	Sac I	Sac II	Sau I	Sca I	Sci I	Sfi I
Sna BI	Spe I	Spi I	Ssp I	Taq IIA	Taq IIA*	Tth 111I
Vsp I	Xca I	Xho I				

Total number of selected enzymes which do not cut: 38

Figure 12a corresponds to the restriction and genetic map of the plasmid pIG2 used to make the intermediary construct pIG2 Mt32 as described in Example IV for the subcloning of the P<sub>32</sub> antigen in plasmid pIGRI.

Figure 12b corresponds to the pIG2 nucleic acid sequence.

On this figure, the origin of nucleotide stretches used to construct plasmid pIG2 is specified hereafter.

Position

3300-206 : lambda PL containing EcoRI-MboII blunt fragment of pPL(λ) (Pharmacia)

207-266 : synthetic sequence containing multiple cloning site and ribosome binding site of which the ATG initiation codon is located at position 232-234

267-772 : rrnBT<sub>1</sub>T<sub>2</sub> containing HindIII-SspI fragment from pKK223 (Pharmacia)

773-3300 : tetracycline resistance gene and origin of replication containing EcoRI-DraI fragment of pAT 153 (Bioexcellence)

Table 7 corresponds to the complete restriction site analysis of pIG2.

Table 7

\*\*\*\*\*  
 \* RESTRICTION-SITE ANALYSIS \*  
 \*\*\*\*\*

Done on DNA sequence PIIG2

Total number of bases is: 3301.  
 Analysis done on the complete sequence.

List of cuts by enzyme.

Acc I	:	252	2647				
AcY I	:	617	2093	2750	2864	2885	
Afl III	:	1527					
Aha III	:	222					
Alu I	:	268	970	1227	1363	1589	2211
Alw NI	:	1118					
Apa LI	:	1213					
Asp 718I	:	208					
Asu I	:	376	505	595	1817	1859	2038
Ava I	:	1872					
Ava II	:	376	1817	1859	2162	2411	2499
Bal I	:	1855					
Bam HI	:	239	2922				
Bbe I	:	2096	2753	2867	2888		
Bbv I	:	271	1198	1617	1635	1748	1751
Bbv I*	:	899	1105	1108	1855	1879	2512
Bbv II	:	1704	2567				
Bgl I	:	2135	2369				
Bin I	:	15	247	785	883	969	2930
Bin I*	:	234	784	881	2195	2917	
Bsp HI	:	737	807	2808			
Bsp MI	:	264	2243				
Bst NI	:	213	357	467	635	1368	1381
Cau II	:	4	565	598	1150	1815	2041
						1502	1857
						2041	2765
						3129	3169

SUBSTITUTE SHEET

Table 7 (Cont'd)

Cfr 10I	:	2014	2368	2528	2887	2896	3137
Cfr I	:	1853	2358	2766	2898	3002	
Cla I	:	3275	262	268	273	303	489
Cvi JI	:	190	1042	1053	1118	1197	1222
Cvi	:	999	1589	1608	1808	1813	1855
		1516	1589	1608	1808	1813	1892
		2182	2192	2211	2252	2309	2317
		2614	2630	2686	2704	2768	2776
		3127	3151	3270	3285		
Cvi QI	:	209	3135	490	546	844	1253
Dde I	:	133	400	779	791	869	877
Dpn I	:	9	241				
Dra II	:	2202	2474	2833	2924	2951	
Dsa I	:	1817	1859	2774			
Eco 3II	:	230	1850	2769			
Eco 47III	:	444	2524	2805	3067		
Eco 57I	:	1655	214				
Eco 57I*	:	985					
Eco 78I	:	2094	2751	2865	2886		
Eco NI	:	196	2674				
Eco RII	:	211	355	465	633	1366	1379
Eco RV	:	3114				1500	1855
Fnu 4HI	:	260	361	913	1119	1122	1187
Fnu	4HI	1737	1740	1869	1883	1890	1893
Fnu DIR	:	2361	2526	2577	2684	2718	2721
Fok I	:	371	903	1484	1666	1763	1885
Fok I*	:	2322	2327	2354	2483	2598	2954
Gsu I	:	297	681	3199			
Gsu I*	:	645	2252	2297	3151		
Gsu I	:	1917					
Gsu I*	:	2471					
Hae I	:	657	673	1053	1505	1855	2252
Hae II	:	423	1287	1657	2096	2526	2753
Hae III	:	507	596	657	673	1053	1487
Hga I	:	2309	2360	2381	2470	2704	2768

SUBSTITUTE SHEET

Table 7. (con't)

Rga I*	:	837	1415	2311	2343	2897	
Rgi AI	:	139	1217	1836	2127	2714	3025
Rgi CI	:	208	2008	2092	2531	2749	2863
Rgi JII	:	2816	2830	422	903	1012	1186
Bha I	:	371	1944	2095	2354	2485	2525
Bin PII	:	369	1942	2093	2352	2483	2523
Hind III	:	3066	3066	3198	2648	107	253
Hind III	:	266	266	3268	1157	1553	1628
Hinf I	:	249	249	564	598	959	1149
Hpa II	:	3	2369	2529	2605	2765	2888
Hph I	:	94	94	138	181	545	796
Hph I*	:	6	212	246	781	1034	1810
Kpn I	:		246	826	1729	1753	2342
Mae I	:		580	991	1107	1170	2149
Mae II	:		169	7	239	777	789
Mae III	:			2200	2472	2831	2922
Mbo I	:			207	304	799	1661
Mbo II	:			870	2826		1709
Mbo II*	:			1134	1318	2994	3081
Mme I*	:			253	1100	1424	1830
Mnl I	:			208	593	1349	1632
Mnl I*	:			3116	3176		1998
Mse I	:			179	186	221	315
Mst I	:			1845	1943	3039	
Nae I	:			2016	2370	2530	2898
Nar I	:			2093	2750	2864	2885
Nco I	:			230			
Nhe I	:			3068			
Nla III	:			166	234	394	449
				2123	2251	2368	2554

SUBSTITUTE SHEET

83 A

		Table 7 (Con't)						
Nla IV	:	210	241	378	1460	1499	1818	1861
	:	2412	2533	2751	2775	2865	2886	2924
Nru I	:	2327						2970
	:		944	1189	2160			
NSP BII	:							
NSP HI	:	1531	2739					
NSP MI	:	1934	1983					
Pfl MI	:							
Ple I	:	257	1636					
Ple I*	:	1151	2660					
Ppu MI	:	1817	1859					
Pss I*	:	1820	1862	2777				
Pst I	:	261						
Rsa I	:	210	3136					
Sal I	:	251	2646					
Sal I	:	4	213	357	467	565	598	635
Scr FI	:							
		1815	1857	2041	2240	2765	3129	3169
Sdu I	:	139	1217	1836	2127	2714	2816	2830
Sec I	:	3	230	1367	1850	1928	2130	3025
Sfa NI	:	479	647	2274	2649	3060	3173	
Sfa NI*	:	1430	1867	2262	2883	2895	3084	
Sph I	:	2739						
Sso II	:	2	211	355	465	563	596	633
		1813	1855	2039	2238	2763	3127	3167
Ssp I	:	226						
Sty I	:	230	1928					
Taq I	:	252	495	1429	2031	2172	2647	2960
Taq IIB	:							
Taq IIB*	:	1631						
		2633						
Tth I	:							
Tth I	:	38	936					
Tth I	:	515	904	943				
Xba I	:	245						
Xba II	:	7	239	777	789	875	886	2922
Xma III	:							
Xma I	:	296						
EcoRI	:	3300						

Total number of cuts is : 689.

SUBSTITUTE SHEET

Table 7 (con't)

## List of non cutting selected enzymes.

Aat	II	,	Afl	II	,	Apa	I	,	Asu	II	,	AVR	II	,	Bbv	II*	,	Bcl	I
Bgl	II	,	Bsp	MI*	,	Bsp	MI*	,	Bss	III	,	Bst	EII	,	Bst	XI	,	Dra	III
Eco	31I*	,	Esp	I	,	Hpa	I	,	Mlu	I	,	Mme	I	,	Nde	I	,	Not	I
Nsi	I	,	Pma	CI	,	Pvu	I	,	Pvu	II	,	Rsr	II	,	Sac	I	,	Sac	II
Sau	I	,	Sca	I	,	Sci	I	,	Sfi	I	,	Sma	I	,	Sma	BI	,	Spe	I
Spl	I	,	Stu	I	,	Taq	IIA	,	Taq	IIA*	,	Tth	111I	,	Vsp	I	,	Xba	I
Xba	I	,	Xma	I															

Total number of selected enzymes which do not cut: 44

Figure 13 corresponds to the amino acid sequence of the total fusion protein mTNF-His<sub>6</sub>-P<sub>32</sub>.

On this figure :

- the continuous underlined sequence (\_\_\_\_) represents the mTNF sequence (first 25 amino acids),
- the dotted underlined sequence (-----) represents the polylinker sequence,
- the double underlined sequence (==) represents the extra amino acids created at cloning site, and
- the amino acid marked with nothing is the antigen sequence starting from the amino acid at position 4 of figure 5.

Figure 14a and 14b correspond to the expression of the mTNF-His<sub>6</sub>-P<sub>32</sub> fusion protein in K12AH, given in Example VI, with Fig. 14a representing the Coomassie Brilliant Blue stained SDS-PAGE and 14b representing immunoblots of the gel with anti-32-kDa and anti-mTNF- antibody.

On fig. 14a, the lanes correspond to the following:

#### Lanes

1.	protein molecular weight markers			
2.	pmTNF-MPH-Mt32	28°C	1 h induction	
3.	"	42°C	"	
4.	"	42°C	2 h induction	
5.	"	42°C	3 h	"
6.	"	28°C	4 h	"
7.	"	42°C	4 h	"
8.	"	28°C	5 h	"
9.	"	42°C	5 h	"

On fig. 14b, the lanes correspond to the following:

Lanes				
1.	pmTNF-MPH-Mt32	28°C	1 h	induction
2.	"	42°C	1 h	"
3.	"	28°C	4 h	"
4.	"	42°C	4 h	"

Figure 15 corresponds to the IMAC elution profile of the recombinant antigen with decreasing pH as presented in Example VII.

Figure 16 corresponds to the IMAC elution profile of the recombinant antigen with increasing imidazole concentrations as presented in Example VII.

Figure 17 corresponds to the IMAC elution profile of the recombinant antigen with a step gradient of increasing imidazole concentrations as presented in Example VII.

#### EXAMPLE I:

#### MATERIAL AND METHODS

##### Screening of the $\lambda$ gt11 *M. tuberculosis* recombinant DNA library with anti-32-kDa antiserum

A  $\lambda$ gt11 recombinant library constructed from genomic DNA of *M. tuberculosis* (Erdman strain), was obtained from R. Young (35). Screening was performed as described (14,35) with some modifications hereafter mentioned.  $\lambda$ gt11 infected *E. coli* Y1090 ( $10^5$  pfu per 150 mm plate) were seeded on NZYM plates (Gibco)(16) and incubated at 42°C for 24 hrs. To induce expression of the  $\beta$ -galactosidase-fusion proteins the plates were overlaid with isopropyl  $\beta$ -D-thiogalactoside (IPTG)-saturated filters (Hybond C extra, Amersham), and incubated for 2 hrs at 37°C. Screening was done with a polyclonal rabbit anti-32-kDa antiserum. Said polyclonal antiserum rabbit anti-32-kDa antiserum was obtained by raising antiserum against the  $P_{32}$  *M. bovis* BCG (strain 1173P2 - Institut Pasteur Paris) as follows: 400  $\mu$ g (purified  $P_{32}$  protein of *M. bovis* BCG) per ml physiological saline were mixed with one volume

of incomplete Freund's adjuvant. The material was homogenized and injected intradermally in 50  $\mu$ l doses, delivered at 10 sites in the back of the rabbits, at 0, 4, 7 and 8 weeks (adjuvant was replaced by the diluent for the last injection). One week later, the rabbits were bled and the sera tested for antibody level before being distributed in aliquots and stored at -80°C.

The polyclonal rabbit anti-32-kDa antiserum was pre-absorbed on *E. coli* lysate (14) and used at a final dilution of 1:300. A secondary alkaline-phosphatase anti-rabbit IgG conjugate (Promega), diluted at 1:5000 was used to detect the  $\beta$ -galactosidase fusion proteins. For color development nitro blue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP) were used. Reactive areas on the filter turned deep purple within 30 min. Usually three consecutive purification steps were performed to obtain pure clones. IPTG, BCIP and NBT were from Promega corp. (Madison WI.).

Plaque screening by hybridization for obtaining the secondary clones BY1, By2 and By5 hereafter defined

The procedure used was as described by Maniatis et al. (14).

Preparation of crude lysates from  $\lambda$ gt11 recombinant lysogens

Colonies of *E. coli* Y1089 were lysogenized with appropriate  $\lambda$ gt11 recombinants as described by Hyunh et al. (14). Single colonies of lysogenized *E. coli* Y1089 were inoculated into LB medium and grown to an optical density of 0.5 at 600nm at 30°C. After a heat shock at 45°C for 20 min., the production of  $\beta$ -galactosidase-fusion protein was induced by the addition of IPTG to a final concentration of 10 mM. Incubation was continued for 60 min. at 37°C and cells were quickly harvested by centrifugation. Cells were concentrated 50 times in buffer (10 mM Tris pH 8.0, 2 mM EDTA) and rapidly frozen into liquid nitrogen. The samples were lysed by

thawing and treated with 100 µg/ml DNase I in EcoRI restriction buffer, for 5-10 minutes at 37°C.

Immunoblotting (Western blotting) analysis:

After SDS-PAGE electrophoresis, recombinant lysogen proteins were blotted onto nitrocellulose membranes (Hybond C, Amersham) as described by Towbin et al. (29). The expression of mycobacterial antigens, fused to  $\beta$ -galactosidase in *E. coli* Y1089 was visualized by the binding of a polyclonal rabbit anti-32-kDa antiserum (1:1000) obtained as described in the above paragraph "Screening of the  $\lambda$ gt11 *M. tuberculosis* recombinant DNA library with anti-32-kDa antiserum" and using a monoclonal anti- $\beta$ -galactosidase antibody (Promega). A secondary alkaline-phosphatase anti-rabbit IgG conjugate (Promega) diluted at 1:5000, was used to detect the fusion proteins.

The use of these various antibodies enables to detect the  $\beta$ -galactosidase fusion protein. This reaction is due to the *M. tuberculosis* protein because of the fact that non fused- $\beta$ -galactosidase is also present on the same gel and is not recognized by the serum from tuberculous patients.

In order to identify selective recognition of recombinant fusion proteins by human tuberculous sera, nitrocellulose sheets were incubated overnight with these sera (1:50) (after blocking aspecific protein binding sites). The human tuberculous sera were selected for their reactivity (high or low) against the purified 32-kDa antigen of *M. bovis* BCG tested in a Dot blot assay as previously described (31). Reactive areas on the nitrocellulose sheets were revealed by incubation with peroxidase conjugated goat anti-human IgG antibody (Dakopatts, Copenhagen, Denmark) (1:200) for 4 hrs and after repeated washings color reaction was developed by adding peroxidase substrate ( $\alpha$ -

chloronaphthol) (Bio-Rad) in the presence of peroxidase and hydrogen peroxide.

#### Recombinant DNA analysis

Initial identification of M. tuberculosis DNA inserts in purified  $\lambda$ gt11 clones was performed by EcoRI restriction. After digestion, the excised inserts were run on agarose gels and submitted to Southern hybridization. Probes were labeled with  $\alpha^{32}P$ -dCTP by random priming (10). Other restriction sites were located by single and double digestions of recombinant  $\lambda$ gt11 phage DNA or their subcloned EcoRI fragments by HindIII, PstI, KpnI, AccI and SphI.

#### Sequencing

Sequence analysis was done by the primer extension dideoxy termination method of Sanger et al. (25) after subcloning of specific fragments in Bluescribe-M13 (6) or in mp10 and mp11 M13 vectors (Methods in Enzymology, vol. 101, 1983, p.20-89, Joachim Messing, New M13 vectors for cloning, Academic Press). Sequence analysis was greatly hampered by the high GC content of the M. tuberculosis DNA (65%). Sequencing reactions were therefore performed with several DNA polymerases: T7 DNA polymerase ("Sequenase" USB), Klenow fragment of DNA polymerase I (Amersham) and in some cases with AMV reverse transcriptase (Super RT, Anglian Biotechnology Ltd.) and sometimes with dITP instead of dGTP. Several oligodeoxynucleotides were synthesized and used to focus ambiguous regions of the sequence. The sequencing strategy is summarized in Fig. 2. In order to trace possible artefactual frameshifts in some ambiguous regions, a special program was used to define the most probable open reading frame in sequences containing a high proportion of GC (3). Several regions particularly prone to sequencing artefacts were confirmed or corrected by chemical sequencing (18). For this purpose, fragments were subcloned in the chemical

sequencing vector pGV462 (21) and analysed as described previously. Selected restriction fragments of about 250-350bp were isolated, made blunt-ended by treatment with either Klenow polymerase or Mung bean nuclease, and subcloned in the SmaI or HincII site of pGV462. Both strands of the inserted DNA were sequenced by single-end labeling at Tth 111I or BstEII (32) and a modified chemical degradation strategy (33).

Routine computer aided analysis of the nucleic acid and deduced amino acid sequences were performed with the LGBC program from Bellon (2). Homology searches used the FASTA programs from Pearson and Lipman (23) and the Protein Identification Resource (PIR) from the National Biomedical Research Fundation - Washington (NBRF) (NBRF/PIR data bank), release 16 (march 1988).

#### RESULTS

##### - Screening of the $\lambda$ gt11M, M. tuberculosis recombinant DNA library with polyclonal anti-32-kDa antiserum :

Ten filters representing  $1.5 \times 10^6$  plaques were probed with a polyclonal rabbit anti-32-kDa antiserum (8). Following purification, six independent positive clones were obtained.

##### Analysis of recombinant clones

EcoRI restriction analysis of these 6 purified  $\lambda$ gt11 recombinant clones DNA, (Fig. 1A) revealed 4 different types of insert. Clone 15 had an insert with a total length of 3.8 kb with two additional internal EcoRI sites resulting in three DNA fragments of 1.8 kb, 1.5 kb and 0.5 kb. The DNA Insert of clone 16 was 1.7 kb long. Clones 17 and 19 had a DNA insert of almost identical length being 2.7 kb and 2.8 kb respectively.

Finally, clone 23 (not shown) and clone 24 both contained an insert of 4 kb with one additional EcoRI restriction site giving two fragments of 2.3 kb and

1.7 kb. Southern analysis (data not shown) showed that the DNA inserts of clones 15, 16, 19 and the small fragment (1.7 kb) of clone 24 only hybridized with themselves whereas clone 17 (2.7 kb) hybridized with itself but also equally well with the 2.3 kb DNA fragment of clone 24. Clones 15, 16 and 19 are thus distinct and unrelated to the 17, 23, 24 group. This interpretation was further confirmed by analysis of crude lysates of *E. coli* Y1089 lysogenized with the appropriate  $\lambda$ gt11 recombinants and induced with IPTG. Western blot analysis (Fig. 1B), showed no fusion protein, either mature or incomplete, reactive with the polyclonal anti-32-kDa antiserum in cells expressing clones 15, 16 and 19. Clones 15, 16 and 19, were thus considered as false positives and were not further studied. On the contrary, cells lysogenized with clone 23 and 24 produced an immunoreactive fusion protein containing about 10 kDa of the 32-kDa protein. Clone 17 finally expressed a fusion protein of which the foreign polypeptide part is about 25 kDa long. The restriction endonuclease maps of the 2.3 kb insert of clone 24 and of the 2.7 kb fragment of clone 17 (Fig. 2) allowed us to align and orient the two inserts suggesting that the latter corresponds to a  $\pm$ 0.5 kb 5' extension of the first.

As clone 17 was incomplete, the same  $\lambda$ gt11 recombinant *M. tuberculosis* DNA library was screened by hybridization with a 120 bp EcoRI-KpnI restriction fragment corresponding to the very 5' end of the DNA insert of clone 17 (previously subcloned in a Blue Scribe vector commercialized by Vector cloning Systems (Stratagene Cloning System) (Fig.2). Three 5'-extended clones By1, By2 and By5 were isolated, analyzed by restriction and aligned. The largest insert, By5 contained the information for the entire coding region

(see below) flanked by 3.1 kb upstream and 1.1 kb downstream (Fig. 2).

#### DNA sequencing

The 1358 base pairs nucleotide sequence derived from the various  $\lambda$ gt11 overlapping clones is represented in Fig. 3a and Fig. 3b. The DNA sequence contains a 1059 base pair open reading frame starting at position 183 and ending with a TAG codon at position 1242. It occurs that the NH<sub>2</sub>-terminal amino-acid sequence, (phe-ser-arg-pro-gly-leu-pro-val-glu-tyr-leu-gln-val-pro-ser-pro-ser-met-gly-arg-asp-ile-lys-val-gln-phe-gln-ser-gly-gly-ala-asn) which can be located within this open reading frame from the nucleotide sequence beginning with a TTT codon at position 360 corresponds to the same NH<sub>2</sub>-terminal amino acid sequence of the MPB 59 antigen except for the amino acids at position 20, 21, 31, which are respectively gly, cys and asn in the MPB 59 (34). Therefore, the DNA region upstream of this sequence is expected to encode a signal peptide required for the secretion of a protein of 32-kDa. The mature protein thus presumably consists of 295 amino acid residues from the N-terminal Phe (TTT codon) to the C-terminal Ala (GCC codon) (Fig. 5).

Six ATG codons were found to precede the TTT at position 360 in the same reading frame. Usage of any of these ATGs in the same reading frame would lead to the synthesis of signal peptides of 29, 42, 47, 49, 55 and 59 residues.

#### Hydropathy pattern

The hydropathy pattern coding sequence of the protein of 32-kDa of the invention and that of the antigen  $\alpha$  of BCG (17) were determined by the method of Kyte and Doolittle (15). The nonapeptide profiles are shown in Fig. 6. Besides the initial hydrophobic signal peptide region, several hydrophilic domains could be

identified. It is interesting to note that the overall hydrophilicity pattern of the protein of 32-kDa of the invention is comparable to that of the BCG antigen  $\alpha$ . For both proteins, a domain of highest hydrophilicity could be identified between amino acid residues 200 and 250.

#### Homology

Matsuo et al. (17) recently published the sequence of a 1095 nucleotide cloned DNA corresponding to the gene coding for the antigen  $\alpha$  of BCG. The 978 bp coding region of *M. bovis* antigen  $\alpha$  as revised in Infection and Immunity, vol. 58, p. 550-556, 1990, and 1017 bp coding regions of the protein of 32-kDa of the invention show a 77.5% homology, in an aligned region of 942 bp. At the amino acid level both precursor protein sequences share 75.6% identical residues. In addition, 17.6% of the amino acids correspond to evolutionary conserved replacements as defined in the algorithm used for the comparison (PAM250 matrix, ref 23). Figure 7 shows sequence divergences in the N-terminal of the signal peptide. The amino terminal sequence - 32 amino acids - of both mature proteins is identical except for position 31.

#### Human sera recognize the recombinant 32-kDa protein

Fig. 8 shows that serum samples from tuberculous patients when immunoblotted with a crude *E. coli* extract expressing clone 17 distinctly react with the 140 kDa fusion protein (lanes 4 to 6) contain the protein of 32-kDa of the invention, but not with unfused  $\beta$ -galactosidase expressed in a parallel extract (lanes 10 to 12). Serum samples from two negative controls selected as responding very little to the purified protein of 32-kDa of the invention does neither recognize the 140 kDa fused protein containing the protein of 32-kDa of the invention, nor the unfused  $\beta$ -galactosidase (lanes 2, 3 and 8 and 9). The 140 kDa  $\beta$ -galactosidase (lanes 1, 5 and 7).

fused protein and the unfused  $\beta$ -galactosidase were easily localized reacting with the anti- $\beta$ -galactosidase monoclonal antibody (lanes 1 to 7).

The invention has enabled to prepare a DNA region coding particularly for a protein of 32-kDa (cf. fig.5); said DNA region containing an open reading frame of 338 codons (stop codon non included). At position 220 a TTT encoding the first amino acid of the mature protein is followed by the 295 triplets coding for the mature protein of 32-kDa. The size of this open reading frame, the immunoreactivity of the derived fusion proteins, the presence of a signal peptide and, especially, the identification within this gene of a  $\text{NH}_2$ -terminal region highly homologous to that found in the MPB 59 antigen (31/32 amino acids homology) and in the BCG antigen  $\alpha$  (31/32 amino acids homology) (see Fig. 7), strongly suggest that the DNA fragment described contains the complete cistron encoding the protein of 32-kDa secreted by M. tuberculosis, and which had never been so far identified in a non ambiguous way.

Six ATG codons were found to precede this TTT at position 220 in the same reading frame. Usage of any of these ATGs in the same reading frame would lead to the synthesis of signal peptides of 43, 48, 50, 56 or 60 residues. Among these various possibilities, initiation is more likely to take place either at ATG<sub>91</sub> or ATG<sub>52</sub> because both are preceded by a plausible E. coli-like promoter and a Shine-Dalgarno motif.

If initiation takes place at ATG<sub>91</sub>, the corresponding signal peptide would code for a rather long peptide signal of 43 residues. This length however is not uncommon among secreted proteins from Gram positive bacteria (5). It would be preceded by a typical E. coli Shine-Dalgarno motif (4/6 residues homologous to AGGAGG) at a suitable distance.

If initiation takes place at ATG<sub>52</sub>, the corresponding signal peptide would code for a peptide signal of 56 residues but would have a less stringent Shine-Dalgarno ribosome binding site sequence.

The region encompassing the translation termination triplet was particularly sensitive to secondary structure effects which lead to so-called compressions on the sequencing gels. In front of the TAG termination codon at position 1105, 22 out of 23 residues are G-C base pairs, of which 9 are G's.

Upstream ATG<sub>130</sub>, a sequence resembling an E. coli promoter (11) comprising an hexanucleotide (TTGAGA) (homology 5/6 to TTGACA) and a AAGAAT box (homology 4/6 to TATAAT) separated by 16 nucleotides was observed. Upstream the potential initiating codon ATG<sub>01</sub>, one could detect several sequences homologous to the E. coli "-35 hexanucleotide box", followed by a sequence resembling a TATAAT box. Among these, the most suggestive is illustrated on Fig. 3a and 3b. It comprises a TTGGCC at position 59 (fig. 3a and 3b) (homology 4/6 to TTGACA) separated by 14 nucleotides from a GATAAG (homology 4/6 to TATAAT). Interestingly this putative promoter region shares no extensive sequence homology with the promoter region described for the BCG protein  $\alpha$ -gene (17) nor with that described for the 65 kDa protein gene (26, 28).

Searching the NBRF data bank (issue 16.0) any significant homology between the protein of 32-kDa of the invention and any other completely known protein sequence could not be detected. In particular no significant homology was observed between the 32-kDa protein and  $\alpha$  and  $\beta$  subunits of the human fibronectin receptor (1). The NH<sub>2</sub>-terminal sequence of the 32-kDa protein of the invention is highly homologous - 29/32 amino acids - to that previously published for BCG-MPB 59 antigen (34) and to that of BCG  $\alpha$ -antigen - 31/32

amino acids - (Matsuo, 17) and is identical in its first 6 amino acids with the 32-kDa protein of M. bovis BCG (9). However, the presumed initiating methionine precedes an additional 29 or 42 amino acid hydrophobic sequence which differs from the one of  $\alpha$ -antigen (cf. Fig. 7), but displaying all the characteristics attributed to signal sequences of secreted polypeptides in prokaryotes (22).

Interestingly, no significant homology between the nucleic acid (1-1358) of the invention (cf. fig. 3a and 3b) and the DNA of the antigen  $\alpha$  of Matsuo exists within their putative promoter regions.

EXAMPLE II: CONSTRUCTION OF A BACTERIAL PLASMID  
CONTAINING THE ENTIRE CODING SEQUENCE OF THE 32-kDa  
PROTEIN OF M. TUBERCULOSIS

In the previous example, in figure 2, the various overlapping  $\lambda$ gt11 isolates covering the 32-kDa protein gene region from *M. tuberculosis* were described. Several DNA fragments were subcloned from these  $\lambda$ gt11 phages in the Blue Scribe M13 $^+$  plasmid (Stratagene). Since none of these plasmids contained the entire coding sequence of the 32-kDa protein gene, a plasmid containing this sequence was reconstructed.

**Step 1 : Preparation of the DNA fragments :**

1) The plasmid BS-By5-800 obtained by subcloning HindIII fragments of By5 (cf. fig. 2) into the Blue Scribe M13 $^+$  plasmid (Stratagene), was digested with HindIII and a fragment of 800 bp was obtained and isolated from a 1% agarose gel by electroelution.

2) The plasmid BS-4.1 obtained by subcloning the 2,7 kb EcoRI insert from  $\lambda$ gt11-17) into the Blue Scribe M13 $^+$  plasmid (Stratagene) (see fig.2 of patent application) was digested with HindIII and SphI and a fragment of 1300 bp was obtained and isolated from a 1% agarose gel by electroelution.

3) Blue Scribe M13' was digested with HindIII and SphI, and treated with calf intestine alkaline phosphatase (special quality for molecular biology, Boehringer Mannheim) as indicated by the manufacturer.

**Step 2 : ligation :**

The ligation reaction contained :

125 ng of the 800 bp HindIII fragment (1)

125 ng of the 1500 bp SphI-HindIII insert (2)

50 ng of the HindIII-SphI digested BSM13' vector

(3)

2  $\mu$ l of 10 ligation buffer (Maniatis et al., 1982)

1  $\mu$ l of (= 2,5 U) of T4 DNA ligase (Amersham)

4  $\mu$ l PEG 6000, 25% (w/v)

8  $\mu$ l H<sub>2</sub>O

The incubation was for 4 hours at 16°C.

**Step 3 : Transformation :**

100  $\mu$ l of DH5 $\alpha$  E. coli (Gibco BRL) were transformed with 10  $\mu$ l of the ligation reaction (step 2) and plated on IPTG, X-Gal ampicillin plates, as indicated by the manufacturer. About 70 white colonies were obtained.

**Step 4 :**

As the 800 bp fragment could have been inserted in both orientations, plasmidic DNA from several clones were analyzed by digestion with PstI in order to select one clone (different from clone 11), characterized by the presence of 2 small fragments of 229 and 294 bp. This construction contains the HindIII-HindIII-SphI complex in the correct orientation. The plasmid containing this new construction was called : "BS.BK.P<sub>32</sub>.complet".

**EXAMPLE III: EXPRESSION OF A POLYPEPTIDE OF THE INVENTION IN E. COLI:**

The DNA sequence coding for a polypeptide, or part of it, can be linked to a ribosome binding site which is part of the expression vector, or can be fused to

the information of another protein or peptide already present on the expression vector.

In the former case the information is expressed as such and hence devoid of any foreign sequences (except maybe for the aminoterminal methionine which is not always removed by E. coli).

In the latter case the expressed protein is a hybrid or a fusion protein.

The gene, coding for the polypeptide, and the expression vector are treated with the appropriate restriction enzyme(s) or manipulated otherwise as to create termini allowing ligation. The resulting recombinant vector is used to transform a host. The transformants are analyzed for the presence and proper orientation of the inserted gene. In addition, the cloning vector may be used to transform other strains of a chosen host. Various methods and materials for preparing recombinant vectors, transforming them to host cells and expressing polypeptides and proteins are described by Panayatatos, N., in "Plasmids, a practical approach (ed. K.G. Hardy, IRL Press) pp.163-176, by Old and Primrose, principals of gene manipulation (2d Ed, 1981) and are well known by those skilled in the art.

Various cloning vectors may be utilized for expression. Although a plasmid is preferable, the vector may be a bacteriophage or cosmid. The vector chosen should be compatible with the host cell chosen.

Moreover, the plasmid should have a phenotypic property that will enable the transformed host cells to be readily identified and separated from those which are not transformed. Such selection genes can be a gene providing resistance to an antibiotic like for instance, tetracyclin, carbenicillin, kanamycin, chloramphenicol, streptomycin, etc.

In order to express the coding sequence of a gene in E. coli the expression vector should also contain

the necessary signals for transcription and translation.

Hence it should contain a promoter, synthetic or derived from a natural source, which is functional in E. coli. Preferably, although usually not absolutely necessary, the promoter should be controllable by the manipulator. Examples of widely used controllable promoters for expression in E. coli are the lac, the trp, the tac and the lambda PL and PR promoter.

Preferably, the expression vector should also contain a terminator of transcription functional in E. coli. Examples of used terminators of transcription are the trp and the rrnB terminators.

Furthermore, the expression vector should contain a ribosome binding site, synthetic or from a natural source, allowing translation and hence expression of a downstream coding sequence. Moreover, when expression devoid of foreign sequences is desired, a unique restriction site, positioned in such a way that it allows ligation of the sequence directly to the initiation codon of the ribosome binding site, should be present.

A suitable plasmid for performing this type of expression is pKK233-2 (Pharmacia). This plasmid contains the trc promoter, the lac Z ribosome binding site and the rrnB transcription terminator.

Also suitable is plasmid pIGRI (Innogenetics, Ghent, Belgium). This plasmid contains the tetracycline resistance gene and the origin of replication of pAT<sub>153</sub> (available from Bioexcellence, Biores B.V., Woerden, The Netherlands), the lambda PL promoter up to the MboII site in the 5' untranslated region of the lambda N gene (originating from pPL(λ); Pharmacia).

Downstream from the PL promoter, a synthetic sequence was introduced which encodes a "two cistron" translation cassette whereby the stop codon of the first

cistron (being the first 25 amino acids of TNF, except for Leu at position 1 which is converted to Val) is situated between the Shine-Dalgarno sequence and the initiation codon of the second ribosome binding site. The restriction and genetic map of pIGRI is represented in Fig. 10a.

Fig. 10b and Table 5 represent respectively the nucleic acid sequence and complete restriction site analysis of pIGRI.

However, when expression as a hybrid protein is desired, then the expression vector should also contain the coding sequence of a peptide or polypeptide which is (preferably highly) expressed by this vector in the appropriate host.

In this case the expression vector should contain a unique cleavage site for one or more restriction endonucleases downstream of the coding sequence.

Plasmids pEX1, 2 and 3 (Boehringer, Mannheim) and pUEX1, 2 and 2 (Amersham) are useful for this purpose.

They contain an ampicillin resistance gene and the origin of replication of pBR322 (Bolivar et al. (1977) Gene 2, 95-113), the lac Z gene fused at its 5' end to the lambda PR promoter together with the coding sequence for the 9 first amino acids of its natural gene cro, and a multiple cloning site at the 3' end of the lac Z coding sequence allowing production of a beta galactosidase fused polypeptide.

The pUEX vectors also contain the CI857 allele of the bacteriophage lambda CI repressor gene.

Also useful is plasmid pmTNF MPH (Innogenetics). It contains the tetracycline resistance gene and the origin of replication of pAT<sub>153</sub> (obtainable from Bioexcellence, Biores B.V., Woerden. The Netherlands), the lambda PL promoter up to the MboII site in the N gene 5' untranslated region (originating from pPL(λ); Pharmacia), followed by a synthetic ribosome binding

100

site (see sequence data), and the information encoding the first 25 AA of mTNF (except for the initial Leu which is converted to Val). This sequence is, in turn, followed by a synthetic polylinker sequence which encodes six consecutive histidines followed by several proteolytic sites (a formic acid, CNBr, kallikrein, and E. coli protease VII sensitive site, respectively), each accessible via a different restriction enzyme which is unique for the plasmid (SmaI, NcoI, BspIII and StuI, respectively; see restriction and genetic map, Fig. 11a). Downstream from the polylinker, several transcription terminators are present including the E. coli trp terminator (synthetic) and the rrnBT<sub>1</sub>T<sub>2</sub> (originating from pKK223-3; Pharmacia). The total nucleic acid sequence of this plasmid is represented in Fig. 11b.

Table 6 gives a complete restriction site analysis of pmTNF MPH.

The presence of 6 successive histidines allows purification of the fusion protein by Immobilized Metal Ion Affinity Chromatography (IMAC).

After purification, the foreign part of the hybrid protein can be removed by a suitable protein cleavage method and the cleaved product can then be separated from the uncleaved molecules using the same IMAC based purification procedure.

In all the above-mentioned plasmids where the lambda PL or PR promoter is used, the promoter is temperature-controlled by means of the expression of the lambda cI ts 857 allele which is either present on a defective prophage incorporated in the chromosome of the host (K12ΔH, ATCC n° 33767) or on a second compatible plasmid (pACYC derivative). Only in the pUEX vectors is this cI allele present on the vector itself.

It is to be understood that the plasmids presented above are exemplary and other plasmids or types of

expression vectors maybe employed without departing from the spirit or scope of the present invention.

If a bacteriophage or phagemid is used, instead of plasmid, it should have substantially the same characteristics used to select a plasmid as described above.

EXAMPLE IV : SUBCLONING OF THE P32 ANTIGEN IN PLASMID pIGRI :

Fifteen  $\mu$ g of plasmid "BS-BK-P<sub>32</sub> complet" (see Example II) was digested with EclXI and BstEII (Boehringer, Mannheim) according to the conditions recommended by the supplier except that at least 3 units of enzyme were used per  $\mu$ g of DNA. EclXI cuts at position 226 (Fig. 5) and BstEII at position 1136, thus approaching very closely the start and stop codon of the mature P<sub>32</sub> antigen. This DNA is hereafter called DNA coding for the "P<sub>32</sub> antigen fragment".

The DNA coding for the "P<sub>32</sub> antigen fragment" (as defined above) is subcloned in pIGRI (see fig. 10a) for expression of a polypeptide devoid of any foreign sequences. To bring the ATG codon of the expression vector in frame with the P<sub>32</sub> reading frame, an intermediary construct is made in pIG2 (for restriction and genetic map, see fig. 12a; DNA sequences, see fig. 12b; complete restriction site analysis, see Table 7).

Five  $\mu$ g of plasmid pIG2 is digested with NcoI. Its 5' sticky ends are filled in prior to dephosphorylation.

Therefore, the DNA was incubated in 40  $\mu$ l NB buffer (0.05 M Tris-Cl pH 7.4; 10 mM MgCl<sub>2</sub>; 0.05%  $\beta$ -mercaptoethanol) containing 0.5 mM of all four dXTP (X = A,T,C,G) and 2  $\mu$ l of Klenow fragment of E. coli DNA polymerase I (5 U/ $\mu$ l, Boehringer, Mannheim) for at least 3 h at 15°C.

After blunting, the DNA was once extracted with one volume of phenol equilibrated against 200 mM Tris-

Cl pH 8, twice with at least two volumes of diethylether and finally collected using the "gene clean kit™" (Bio101) as recommended by the supplier. The DNA was then dephosphorylated at the 5' ends in 30  $\mu$ l of CIP buffer (50 mM TrisCl pH 8, 1 mM ZnCl<sub>2</sub>) and 20 to 25 units of calf intestine phosphatase (high concentration, Boehringer, Mannheim). The mixture was incubated at 37°C for 30 min, then EGTA (ethyleneglycol bis ( $\beta$ -aminoethylether)-N,N,N',N' tetraacetic acid) pH 8 is added to a final concentration of 10 mM. The mixture was then extracted with phenol followed by diethylether as described above, and the DNA was precipitated by addition of 1/10 volume of 3 M KAC (Ac = CH<sub>3</sub>COO) pH 4.8 and 2 volumes of ethanol followed by storage at -20°C for at least one hour.

After centrifugation at 13000 rpm in a Biofuge A (Heraeus) for 5 min the pelleted DNA was dissolved in H<sub>2</sub>O to a final concentration of 0.2  $\mu$ g/ $\mu$ l.

The EclXI-BstEII fragment, coding for the "P<sub>32</sub> antigen fragment" (see above) was electrophoresed on a 1% agarose gel (BRL) to separate it from the rest of the plasmid and was isolated from the gel by centrifugation over a Millipore HVLP filter ( $\phi$  2 cm) (2 min., 13000 rpm, Biofuge at room temperature) and extracted with Tris equilibrated phenol followed by diethylether as described above.

The DNA was subsequently collected using the "Gene clean kit™" (Bio101) as recommended by the supplier.

After that, the 5' sticky ends were blunted by treatment with the Klenow fragment of E. coli DNA polymerase I as described above and the DNA was then again collected using the "Gene clean kit™" in order to dissolve it in 7  $\mu$ l of H<sub>2</sub>O.

One  $\mu$ l of vector DNA is added together with one  $\mu$ l of 10 x ligase buffer (0.5 M TrisCl pH 7.4, 100 mM MgCl<sub>2</sub>, 5 mM ATP, 50 mM DTT (dithiothreitol)) and 1  $\mu$ l

of T4 DNA ligase (1 unit/ $\mu$ l, Boehringer, Mannheim). Ligation was performed for 6 h at 13°C and 5  $\mu$ l of the mixture is then used to transform strain DH1 (lambda) [strain DH1 - ATCC N° 33849 - lysogenized with wild type bacteriophage  $\lambda$ ] using standard transformation techniques as described for instance by Maniatis et al. in "Molecular cloning, a laboratory manual", Cold Spring Harbor Laboratory (1982).

Individual transformants are grown and lysed for plasmid DNA preparation using standard procedures (Experiments with gene fusions, Cold Spring Harbor Laboratory (1984) (T.J. Silhavy, H.L. Berman and L.W. Enquist, eds) and the DNA preparations are checked for the correct orientation of the gene within the plasmid by restriction enzyme analysis.

A check for correct blunting is done by verifying the restoration of the NcoI site at the 5' and 3' end of the antigen coding sequence. One of the clones containing the  $P_{32}$  antigen fragment in the correct orientation is kept for further work and designated pIG<sub>2</sub>-Mt32. In this intermediary construct, the DNA encoding the antigen is not in frame with the ATG codon. However, it can now be moved as a NcoI fragment to another expression vector.

15  $\mu$ g of pIG<sub>2</sub>-Mt32 is digested with NcoI. The NcoI fragment encoding the  $P_{32}$  antigen is gel purified and blunted as described above. After purification, using "gene clear kit TM" it is dissolved in 7  $\mu$ l of H<sub>2</sub>O.

5  $\mu$ g of plasmid pIGRI is digested with NcoI, blunted and dephosphorylated as described above. After phenol extraction, followed by diethylether and ethanolprecipitation, the pellet is dissolved in H<sub>2</sub>O to a final concentration of 0.2  $\mu$ g/ $\mu$ l.

Ligation of vector and "antigen fragment" DNA is carried out as described above. The ligation mixture is then transformed into strain DH1 (lambda) and

individual transformants are analysed for the correct orientation of the gene within the plasmid by restriction enzyme analysis. A check for correct blunting is done by verifying the creation of a new NsiI site at the 5' and 3' ends of the antigen coding sequence. One of the clones containing the  $P_{32}$  antigen fragment in the correct orientation is kept for further work and designated pIGRI.Mt32.

EXAMPLE V: SUBCLONING OF THE "P<sub>32</sub> ANTIGEN IN pmTNF MPH:

Fifteen  $\mu$ g of the plasmid pIG2 Mt32- (see example IV) was digested with the restriction enzyme NcoI (Boehringer, Mannheim), according to the conditions recommended by the supplier except that at least 3 units of enzyme were used per  $\mu$ g of DNA.

After digestion, the reaction mixture is extracted with phenol equilibrated against 200mM TrisCl pH 8, (one volume), twice with diethylether (2 volumes) and precipitated by addition of 1/10 volume of 3 M KAc ( $Ac=CH_3COO$ ) pH 4.8 and 2 volumes of ethanol followed by storage at -20°C for at least one hour.

After centrifugation for 5 minutes at 13000 rpm in a Biofuge A (Hereaus) the DNA is electrophoresed on a 1% agarose gel (BRL).

The DNA coding for the "P<sub>32</sub> antigen fragment" as described above, is isolated by centrifugation over a Millipore HVLP filter ( $\phi$  2cm)(2 minutes, 13000 rpm, Biofuge at room temperature) and extracted one with trisCl equilibrated phenol and twice with diethylether. The DNA is subsequently collected using "Gene clean kit 1.M." (Bio 101) and dissolved in 7 $\mu$ l of H<sub>2</sub>O.

The 5' overhanging ends of the DNA fragment generated by digestion with NcoI were filled in by incubating the DNA in 40  $\mu$ l NB buffer (0.05 M Tris-HCl, pH 7.4; 10 mM MgCl<sub>2</sub>; 0.05%  $\beta$ -mercaptoethanol) containing 0.5 mM of all four dXTPS (X = A, T, C, G) and 2 $\mu$ l of Klenow fragment of E. coli DNA polymerase I

(5 units/ $\mu$ l Boehringer Mannheim) for at least 3 h at 15°C. After blunting, the DNA was extracted with phenol, followed by diethylether, and collected using a "gene clean kit" as described above.

Five  $\mu$ g of plasmid pmTNF MPH is digested with StuI, subsequently extracted with phenol, followed by diethylether, and precipitated as described above. The restriction digest is verified by electrophoresis of a 0.5  $\mu$ g sample on an analytical 1,2% agarose gel.

The plasmid DNA is then desphosphorylated at the 5' ends to prevent self-ligation in 30 $\mu$ l of CIP buffer (50 mM TrisCl pH 8, 1 mM ZnCl<sub>2</sub>) and 20 to 25 units of calf intestine phosphatase (high concentration, Boehringer Mannheim). The mixture is incubated at 37°C for 30 minutes, then EGTA (ethyleneglycol bis ( $\beta$ -aminoethylether)-N,N,N',N' tetraacetic acid) pH8 is added to a final concentration of 10 mM. The mixture is extracted with phenol followed by diethylether and the DNA is precipitated as described above. The precipitate is pelleted by centrifugation in a Biofuge A (Heraeus) at 13000 rpm for 10 min at 4°C and the pellet is dissolved in H<sub>2</sub>O to a final DNA concentration of 0.2  $\mu$ g/ $\mu$ l.

One  $\mu$ l of this vector DNA is mixed with the 7  $\mu$ l solution containing the DNA fragment coding for the "P32antigen fragment" (as defined above) and 1  $\mu$ l 10 x ligase buffer (0.5 M TrisCl pH7.4, 100 mM MgCl<sub>2</sub>, 5 mM ATP, 50 mM DTT (dithiothreitol)) plus 1  $\mu$ l T<sub>4</sub> DNA ligase (1 unit/ $\mu$ l, Boehringer Mannheim) is added. The mixture is incubated at 13°C for 6 hours and 5  $\mu$ l of the mixture is then used for transformation into strain DH1(lambda) using standard transformation techniques are described by for instance Maniatis et al. in "Molecular cloning, a laboratory manual", Cold Spring Harbor Laboratory (1982).

Individual transformants are grown and then lysed for plasmid DNA preparation using standard procedures

(Experiments with gene fusions, Cold Spring Harbor Laboratory 1984 (T.J. Silhavy, M.L. Berman and L.W. Enquist eds.)) and are checked for the correct orientation of the gene within the plasmid by restriction enzyme analysis.

One of the clones containing the DNA sequence encoding the antigen fragment in the correct orientation was retained for further work and designated pmTNF-MPH-Mt32. It encodes all information of the  $P_{32}$  antigen starting from position +4 in the amino acid sequence (see fig. 5). The amino acid sequence of the total fusion protein is represented in fig. 13.

EXAMPLE VI: INDUCTION OF ANTIGEN EXPRESSION FROM pmTNF MPH Mt32 :

A- MATERIAL AND METHODS

DNA of pmTNF-MPH-Mt32 is transformed into E. coli strain K12ΔH (ATCC 33767) using standard transformation procedures except that the growth temperature of the cultures is reduced to 28°C and the heat shock temperature to 34°C.

A culture of K12ΔH harboring pmTNF-MPH-Mt32, grown overnight in Luria broth at 28°C with vigorous shaking in the presence of 10 µg/ml tetracycline, is inoculated into fresh Luria broth containing tetracycline (10 µg/ml) and grown to an optical density at 600 nanometers of 0.2 in the same conditions as for the overnight culture.

When the optical density at 600 nanometers has reached 0.2 half of the culture is shifted to 42°C to induce expression while the other half remains at 28°C as a control. At several time intervals aliquots are taken which are extracted with one volume of phenol equilibrated against M9 salts (0.1% ammonium chloride, 0.3% potassium dihydrogenium phosphate, 1.5% disodium hydrogenium phosphate, 12 molecules of water) and 1% SDS. At the same time, the optical density (600 nm) of

the culture is checked. The proteins are precipitated from the phenol phase by addition of two volumes of acetone and storage overnight at -20°C. The precipitate is pelleted (Biofuge A, 5 min., 13000 rpm, room temperature) dried at the air, dissolved in a volume of Laemmli (Nature (1970) 227:680) sample buffer (+  $\beta$  mercapto ethanol) according to the optical density and boiled for 3 min.

Samples are then run on a SDS polyacrylamide gel (15%) according to Laemmli (1970). Temperature induction of mTNF-His<sub>6</sub>-P<sub>32</sub> is monitored by both Coomassie Brilliant Blue (CBB) staining and immunoblotting. CBB staining is performed by immersing the gel in a 1/10 diluted CBB staining solution (0.5 g CBB-R250 (Serva) in 90 ml methanol : H<sub>2</sub>O (1:1 v/v) and 10 ml glacial acetic acid) and left for about one hour on a gently rotating platform. After destaining for a few hours in destaining solution (30% methanol, 7% glacial acetic acid) protein bands are visualised and can be scanned with a densitometer (Ultroscan XL Enhanced Laser Densitometer, LKB).

For immunoblotting the proteins are blotted onto Hybond C membranes (Amersham) as described by Townbin et al (1979). After blotting, proteins on the membrane are temporarily visualised with Ponceau S (Serva) and the position of the molecular weight markers is indicated. The stain is then removed by washing in H<sub>2</sub>O. Aspecific protein binding sites are blocked by incubating the blots in 10% non-fat dried milk for about 1 hour on a gently rotating platform. After washing twice with NT buffer (25 mM Tris-HCl, pH 8.0; 150 mM NaCl) blots are incubated with polyclonal rabbit anti-32-kDa antiserum (1:1000), obtained as described in example I ("screening of the  $\lambda$ gt11 M. tuberculosis recombinant DNA library with anti-32-kDa antiserum") in the presence of E. coli lysate or with monoclonal

anti-hTNF-antibody which crossreacts with mTNF (Innogenetics, n° 17F5D10) for at least 2 hours on a rotating platform. After washing twice with NT buffer + 0.02% Triton X.100, blots are incubated for at least 1 hour with the secondary antiserum : alkaline phosphatase-conjugated swine anti-rabbit immunoglobulins (1/500; Prosan) in the first case, and alkaline phosphatase conjugated rabbit anti-mouse immunoglobulins (1/500; Sigma) in the second case.

Blots are washed again twice with NT buffer + 0.02% Triton X100 and visualisation is then performed with nitro blue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) from Promega using conditions recommended by the supplier.

#### B. RESULTS

Upon induction of K12ΔH cells containing pmTNF-MPH-Mt32, a clearly visible band of about 35-kDa appears on CBB stained gels, already one hour after start of induction (Fig. 14a). This band, corresponding to roughly 25% of total protein contents of the cell, reacts strongly with anti-32-kDa and anti-mTNF antisera on immunoblots (Fig. 14b). However, this band represents a cleavage product of the original fusion protein, since a minor band, around 37 kDa, is also visible on immunoblots, reacting specifically with both antisera as well. This suggests that extensive cleavage of the recombinant mTNF-His<sub>6</sub>-P<sub>32</sub> takes place about 2-3 kDa from its carboxyterminal end.

#### EXAMPLE VII : PURIFICATION OF RECOMBINANT ANTIGEN ON IMMOBILIZED METAL ION AFFINITY CHROMATOGRAPHY (IMAC) :

The hybrid protein mTNF-His<sub>6</sub>-P<sub>32</sub> (amino acid sequence, see fig. 13) expressed by K12ΔH cells containing pmTNF.MPH.Mt32, is especially designed to facilitate purification by IMAC, since the 6 successive histidines in the polylinker sequence bring about a strong affinity for metal ions (HOCHULI et al, 1988).

a. Preparation of the crude cell extract :

12 l of E. coli cells K12ΔH containing plasmid pmTNF-MPH-Mt32 were grown in Luria Broth containing tetracycline (10 µg/ml) at 28°C to an optical density (600 nm) of 0.2 and then induced by shifting the temperature to 42°C. After 3 hours of induction, cells were harvested by centrifugation (Beckman, JA 10 rotor, 7.500 rpm, 10 min). The cell paste was resuspended in lysis buffer (10 mM KCl, 10 mM Tris-HCl pH 6.8, 5 mM EDTA) to a final concentration of 50% (w/v) cells.

$\epsilon$ -NH<sub>2</sub>-capronic acid and dithiotreitol (DTT) were added to a final concentration of resp. 20 mM and 1 mM, to prevent proteolytic degradation. This concentrated cell suspension was stored overnight at -70°C.

Cells were lysed by passing them three times through a French press (SLM-Aminco) at a working pressure of 800-1000 psi. During and after lysis, cells were kept systematically on ice.

The cell lysate was cleared by centrifugation (Beckman, JA 20, 18.000 rpm, 20 min, 4°C). The supernatant (SN) was carefully taken off and the pellet, containing membranes and inclusion bodies, was kept for further work since preliminary experiments had shown that the protein was mainly localised in the membrane fraction.

7 M guanidinium hydrochloride (GuHCl, marketed by ICN) in 100 mM phosphate buffer pH 7.2 was added to the pellet volume to a final concentration of 6 M GuHCl. The pellet was resuspended and extracted in a bounce tissue homogenizer (10 cycles).

After clearing (Beckman, JA 20, 18.000 rpm, 20 min, 4°C), about 100 ml of supernatant was collected (= extract 1) and the removing pellet was extracted again as described above (= extract 2, 40 ml).

The different fractions (SN, EX1, EX2) were analysed on SDS-PAGE (Laemmli, Nature 1970; 227:680) together

with control samples of the induced culture. Scanning of the gel revealed that the recombinant protein makes up roughly 25% of the total protein content of the induced cell culture. After fractionation most of the protein was found back in the extracts. No difference was noticed between reducing and non-reducing conditions (plus and minus  $\beta$ -mercaptoethanol).

b. Preparation of the  $\text{Ni}^{++}$  IDA (Imino diacetic acid) column :

5 ml of the chelating gel, Chelating Sepharose 6B (Pharmacia) is washed extensively with water to remove the ethanol in which it is stored and then packed in a "Econo-column" (1 x 10 cm, Biorad). The top of the column is connected with the incoming fluid (sample, buffer, etc) while the end goes to the  $\text{UV}_{280}$  detector via a peristaltic pump. Fractions are collected using a fraction collector and, when appropriate, pH of the fractions is measured manually.

The column is loaded with  $\text{Ni}^{++}$  (6 ml  $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$ ; 5  $\mu\text{g}/\mu\text{l}$ ) and equilibrated with starting buffer (6 M guanidinium hydrochloride, 100 mM phosphate buffer, pH 7.2).

After having applied the sample, the column is washed extensively with starting buffer to remove unbound material.

To elute the bound material, 2 different elution procedures are feasible :

- 1) elution by decreasing pH,
- 2) elution by increasing imidazol concentration.

Both will be discussed here.

To regenerate the column, which has to be done after every 2-3 runs, 20 ml (about 5 column volumes) of the following solutions are pumped successively through the column :

- 0.05 M EDTA - 0.5 M NaCl
- 0.1 M NaOH

- $\text{H}_2\text{O}$
- 6 ml  $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$  (5 mg/ml).

After equilibrating with starting buffer the column is ready to use again.

c. Chromatography :

All buffers contained 6 M guanidinium hydrochloride throughout the chromatography. The column was developed at a flow rate of 0.5 ml/min at ambient temperature. Fractions of 2 ml were collected and, when appropriate, further analysed by SDS-PAGE and immunoblotting. Gels were stained with Coomassie Brilliant Blue R250 and silver stain, as described by ANSORGE (1985). Immunoblotting was carried out as described in example I. The primary antiserum used was either polyclonal anti-32kDa-antiserum (1/1000) obtained as described in example I ("screening of the  $\lambda$ gt11 M. tuberculosis recombinant DNA library with anti-32kDa-antiserum") or anti-E. coli-immunoglobulins (1/500; PROSAN), or monoclonal anti-hTNF-antibody which cross-reacts with mTNF (Innogenetics, N° 17F5D10). The secondary antiserum was alkaline phosphatase conjugated swine anti-rabbit immunoglobulins (1/500, PROSAN), or alkaline phosphatase conjugated rabbit-anti-mouse immunoglobulins (1/500, Sigma).

C1. Elution with decreasing pH :

Solutions used :

A : 6 M GuHCl 100 mM phosphate pH 7.2

B : 6 M GuHCl 25 mM phosphate pH 7.2

C : 6 M GuHCl 50 mM phosphate pH 4.2

After applying 3 ml of extract 1 ( $\text{OD}_{280} = 32.0$ ) and extensively washing with solution A, the column is equilibrated with solution B and then developed with a linear pH gradient from 7.2 to 4.2 (25 ml of solution B and 25 ml of solution C were mixed in a gradient former). The elution profile is shown in figure 15.

From SDS-PAGE analysis (Coomassie and silverstain) it was clear that most of the originally bound recombinant protein was eluted in the fractions between pH 5.3 and 4.7.

Screening of these fractions on immunoblot with anti-32-kDa and the 17F5D10 monoclonal antibody showed that, together with the intact recombinant protein, also some degradation products and higher aggregation forms of the protein were present, although in much lower amount. Blotting with anti-E. coli antibody revealed that these fractions (pH 5.3-4.7) still contained immunodetectable contaminating E. coli proteins (75, 65, 43, 35 and 31 kDa bands) and lipopolysaccharides..

c2. Elution with increasing imidazol concentration :

Solutions used :

A : 6 M GuHCl 100 mM phosphate pH 7.2

B : 6 M GuHCl 50 mM imidazol pH 7.2

C : 6 M GuHCl 100 mM imidazol pH 7.2

D : 6 M GuHCl 15 mM imidazol pH 7.2

E : 6 M GuHCl 25 mM imidazol pH 7.2

F : 6 M GuHCl 35 mM imidazol pH 7.2

Sample application and washing was carried out as in C1, except that after washing, no equilibration was necessary with 6 M GuHCl 25 mM phosphate. The column was first developed with a linear gradient of imidazol going from 0 to 50 mM (25 ml of solution A and 25 ml of solution B were mixed in a gradient former) followed by a step elution to 100 mM imidazol (solution C). During the linear gradient, proteins were gradually eluted in a broad smear, while the step to 100 mM gave rise to a clear peak (fig. 16).

SDS-PAGE analysis of the fractions revealed that in the first part of the linear gradient (fr 1-24) most

contaminating E. coli proteins were washed out, while the latter part of the gradient (fr 25-50) and the 100 mM peak contained more than 90% of the recombinant protein.

As in C1, these fractions showed, besides a major band of intact recombinant protein, some minor bands of degradation and aggregation products. However, in this case, the region below 24-kDa seemed nearly devoid of protein bands, which suggests that less degradation products co-elute with the intact protein. Also, the same contaminating E. coli proteins were detected by immunoblotting, as in C1, although the 31-kDa band seems less intense and even absent in some fractions.

In a second stage, we developed the column with a step gradient of increasing imidazol concentrations. After having applied the sample and washed the column, 2 column volumes (about 8 ml) of the following solutions were brought successively onto the column : solution D, E, F and finally 4 column volumes of solution C. The stepgradient resulted in a more concentrated elution profile (fig. 17) which makes it more suitable for scaling up purposes.

In conclusion, the mTNF-His<sub>6</sub>-P<sub>32</sub> protein has been purified to at least 90% by IMAC. Further purification can be achieved through a combination of the following purification steps :

- IMAC on chelating superose (Pharmacia)
- ion exchange chromatography (anion or cation)
- reversed phase chromatography
- gel filtration chromatography
- immunoaffinity chromatography
- elution from polyacrylamide gel.

These chromatographic methods are commonly used for protein purification.

The plasmids of figures 10b, 11b and 12b are new.

## BIBLIOGRAPHY

1. Abou-Zeid, C., T.L. Ratliff, H.G. Wiker, M. Harboe, J. Bennedsen and G.A.W. Rook, 1988. Characterization of fibronectin-biding antigens released by *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG. *Infect. Imm.* 56, 3046-3051.
2. Bellon, B. 1988. Apple Macintosh programs for nucleic and protein sequence analysis. *Nucleic Acid Res.* 16:1837-1846.
3. Bibb, M.J., P.R. Findlay and M.W. Jonhson. 1984. The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. *Gene.* 30: 157-166.
4. Bresson, G.M. and K.K. Stanley. 1987. pUEX, a bacterial expression vector related to pEX with universal host specificity. *Nucl. Aci. Res.* 15:10056.
5. Chang, S. Engineering for protein secretion in Gram positive bacteria. *Methods Enzymol.*, 153:507-516.
6. Chen, E.J. and P.H. Seeburg. 1985. Supercoil sequencing: a fast simple method for sequencing plasmid DNA. *DNA* 4:165-170.
7. Closs, O., M. Harboe, N.H. Axelsen-Christensen and M. Magnussen. 1980. The antigens of *Mycobacterium bovis*, strain BCG, studied by cross-immuno-electrophoresis: a reference system. *Scand. J. Immunol.* S12N:249-263.
8. De Bruyn, J.R. Bosmans, J. Nyabenda and J.P. Van Vooren. 1989. Effect of zinc deficiency of the appearance of two immunodominant protein antigens (32-kDa and 65-kDa) in culture filtrates of *Mycobacteria*. *J. Gen. Micriob.* 135: 79-84.
9. De Bruyn, J., K. Huygen, R. Bosmans, M. Fauville, R. Lippens, J.P. Van Vooren, P. Falmagne, M. Weckx, H.G. Wiker, M. Harboe and M. Turneer. 1987. Purification,

partial characterization and identification of a 32-kDa protein antigen of *Mycobacterium bovis* BCG. *Microb. Pathogen.* 2:351-366.

10. Felnberg, A.P. and R. Vogelstein. 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6-13.

11. Hawley, D.K. and W.R. Mc Clure. 1983. Compilation and analysis of *E. coli* promoter DNA sequences. *Nucleic Acids Res.* 11:2237-2255.

12. Huygen, K., J.P. Van Vooren, M. Turneer, R. Bosmans, P. Dierckx and J. De Bruyn. 1988. Specific lymphoproliferation -interferon production and serum immunoglobulin G directed against a purified 32-kDa Mycobacterial antigen (P32) in patient with active tuberculosis. *Scand. J. Immunol.* 27:187-194.

13. Huygen, K., K. Palfliet, F. Jurton, J. Hilgers, R. ten Berg, J.P. Van Vooren and J. De Bruyn. 1989. H-2-linked control of in vitro interferon production in response to 32-kilodalton (P32) of *Mycobacterium bovis* bacillus Calmette-Guérin. *Infect. Imm.* 56:3196-3200.

14. Huynh, T.V., R.A. Young and R.W. Davis. 1985. Constructing and screening libraries in gt10 and gt11. p.49-78. in: *DNA cloning*. Vol.I, A practical approach. Ed. D.M. Glover. IRL Press, Oxford-Washington, D.C.

15. Kyte, J. and R.F. Doolittle. 1982. Simple method for displaying the hydropathy character of a protein. *J. Mol. Biol.* 157:105-132.

16. Maniatis, T., E.F. Fritsch and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

17. Matsuo, K., R. Yamaguchi, A. Yamazaki, H. Tasaka and T. Yamada. 1988. Cloning and expression of the *Mycobacterium bovis* BCG gene for extracellular  $\alpha$ -antigen. *J. Bacteriol.* 170:3847-3854.

18. Mawam, A.M. and W. Gilbert. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA.* 74:560-564.
19. Mehra, V., D. sweetser and R.A. Young. 1986. Efficient mapping of protein antigenic determinants. *Proc. Natl. Acad. Sci. USA.* 83:7013-7017.
20. Mustafa, A.B., H.K. Gill, A. Nerland, W.J. Britton, V. Mehra, B.R. Bloom, R.A. Young and T. Godal. 1986. Human T-cell clones recognize a major *M.Leprae* protein antigen expressed in *E. coli*. *Nature (London)*. 319:63-38.
21. Neesen, K. and G. Volckaert. 1989. Construction and shuttling of novel bifunctional vectors for *Streptomyces* spp. and *Escherichia coli*. *J. Bacteriol.* 171:1569-1573.
22. Oliver, D. 1985. Protein secretion in *Escherichia coli*. *Ann. Rev. Microbiol.* 39:615-648.
23. Pearson, W.R. and D.J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA.* 85:2444-2448.
24. Rumschlag, H.S., T.S. Shinnick and M.L. Cohen. 1988. Serological response of patients with lepromatous and tuberculous leprosy to 30-, 31- and 32-kilodalton antigens of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 26:2200-2202.
25. Sanger, F., S. Niklon and A.R. Coulson. 1977. DNA sequencing with chain termination inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463-5487.
26. Shinnick, T.M. 1987. The 65-kilodalton antigen of *Mycobacterium tuberculosis*. *J. Bacteriol.* 169:1080-1088.
27. Thole, J.E.R., W.C.A. Van Shooten, W.J. Keulen, P.W.M. Hermans, A.A., M. Janson, R.R.P. De Vries, A.H.J. Kolk and J.D.A. Van Embden. 1988. Use of recombinant antigens expressed in *Escherichia coli* K-12 to map B-cell and T-cell epitopes on the immunodominant

65-kilodalton protein of *Mycobacterium bovis* BCG. *Infect. Immun.* 56:1633-1640.

28. Thole. J.E.R., W.J. Keulen, J. De Bruyn, A.H.J. Kolk, D.G. Groothuis, L.G. Berwald, R.H. Tiesjema and J.D.A. Van Embden. 1987. Characterization, sequence determination and immunogenicity of a 64-kilodalton protein of *Mycobacterium bovis* BCG expressed in *Escherichia coli* K-12. *Infect. Immun.* 55:1466-1475.

29. Towbin, H., T. Staehelin and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. USA* 76:4350-4354.

30. Turneer, M., J.P. Van Vooren, J. De Bruyn, E. Serruys, P. Dierckx and J.C. Yernault. 1988. Humoral immune response in human tuberculosis: immunoglobulins G, A and M directed against the purified P32 protein antigen of *Mycobacterium bovis* bacillus Calmette-Guérin. *J. Clin. Microbiol.* 26:1714-1719.

31. Van Vooren, J.P., C.M. Farber, E. Noël, N. Mavroudakis, M. Turneer, J. De Bruyn, F. Legros and J.C. Yernault. 1989 Local anti-P32 humoral response in tuberculous meningitis. *Tubercle*. 70:123-126.

32. Volckaert, G. 1987. A systematic approach to chemical sequencing by subcloning in pGV451 and derived vectors. *Methods Enzymol.* 155:231-250.

33. Volckaert, G., El. De Vieeschouwer, R. Frank and H. Bloecker. 1984. A novel type of cloning vectors for ultrarapid chemical degradation sequencing of DNA. *Gene Anal. Techn.* 1:52-59.

34. Wiker, H.G., M. Harboe, S. Nagal, M.E. Patarroyo, C. Ramirez and N. Cruz. 1986. MPB59, a widely cross-reacting protein of *Mycobacterium bovis* BCG. *Int. Arch. Allergy Appl. Immunol.* 81:307-314.

35. Young, R.A., B.R. Bloom, C.M. Grosskinsky, J. Ivanji, D. Thomas and R.W. Davis. 1985. Dissection of

Mycobacterium tuberculosis antigens using recombinant DNA. Proc. Natl. Acad. Sci. USA, 82:2583-2587.

36. HOCHULI, E., BANNWARTH, W., DÖBELI, H., GENTZ, R. and STÜBER, D. (1988). Genetic Approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. Biotechnology, nov. 1988, p. 1321-1325.

37. ANSORGE, W. (1985), Fast and sensitive detection of protein and DNA bands by treatment with potassium permanganate. J. Biochem. Biophys. Meth., 11:13-20.

CLAIMS

1. Recombinant polypeptide containing in its polypeptidic chain, one at least of the following amino acid sequences:
  - the one extending from the extremity constituted by amino acid at position (-29) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b, or
  - the one extending from the extremity constituted by amino acid at position (12) to the extremity constituted by amino acid at position (31) represented on fig. 3a and fig. 3b, or
  - the one extending from the extremity constituted by amino acid at position (36) to the extremity constituted by amino acid at position (55) represented on fig. 3a and fig. 3b, or
  - the one extending from the extremity constituted by amino acid at position (77) to the extremity constituted by amino acid at position (96) represented on fig. 3a and fig. 3b, or
  - the one extending from the extremity constituted by amino acid at position (101) to the extremity constituted by amino acid at position (120) represented on fig. 3a and fig. 3b, or
  - the one extending from the extremity constituted by amino acid at position (175) to the extremity constituted by amino acid at position (194) represented on fig. 3a and fig. 3b, or
  - the one extending from the extremity constituted by amino acid at position (211) to the extremity constituted by amino acid at position (230) represented on fig. 3a and fig. 3b, or
  - the one extending from the extremity constituted by amino acid at position (275) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,

120

and the peptidic sequences resulting from the modification by substitution and/or by addition and/or by deletion of one or several amino acids in so far as this modification does not alter the following properties :

the polypeptides react with rabbit polyclonal antiserum raised against the protein of 32-kDa of M. bovis BCG culture filtrate, and/or

react selectively with human sera from tuberculosis patients and particularly patients developing an evolutive tuberculosis at an early stage, and/or react with the amino acid sequence extending from the extremity constituted by amino acid at position (1), to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b.

2. Recombinant polypeptide according to claim 1, containing in its polypeptidic chain, one at least of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (-29) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b, or

- the one extending from the extremity constituted by amino acid at position (12) to the extremity constituted by amino acid at position (31) represented on fig. 4a and fig. 4b, or

- the one extending from the extremity constituted by amino acid at position (36) to the extremity constituted by amino acid at position (55) represented on fig. 4a and fig. 4b, or

- the one extending from the extremity constituted by amino acid at position (77) to the extremity constituted by amino acid at position (96) represented on fig. 4a and fig. 4b, or

- the one extending from the extremity constituted by amino acid at position (101) to the extremity constituted by amino acid at position (120) represented on fig. 4a and fig. 4b, or
- the one extending from the extremity constituted by amino acid at position (175) to the extremity constituted by amino acid at position (194) represented on fig. 4a and fig. 4b, or
- the one extending from the extremity constituted by amino acid at position (211) to the extremity constituted by amino acid at position (230) represented on fig. 4a and fig. 4b, or
- the one extending from the extremity constituted by amino acid at position (275) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,

and the peptidic sequences resulting from the modification by substitution and/or by addition and/or by deletion of one or several amino acids in so far as this modification does not alter the following properties :

the polypeptides react with rabbit polyclonal antiserum raised against the protein of 32-kDa of M. bovis BCG culture filtrate, and/or

react selectively with human sera from tuberculosis patients and particularly patients developing an evolutive tuberculosis at an early stage, and/or react with the amino acid sequence extending from the extremity constituted by amino acid at position (1), to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b.

3. Recombinant polypeptide according to claim 1, containing in its polypeptidic chain, one at least of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (-30) to the extremity constituted by amino acid at position (-1) represented on fig. 5, or
- the one extending from the extremity constituted by amino acid at position (12) to the extremity constituted by amino acid at position (31) represented on fig. 5, or
- the one extending from the extremity constituted by amino acid at position (36) to the extremity constituted by amino acid at position (55) represented on fig. 5, or
- the one extending from the extremity constituted by amino acid at position (77) to the extremity constituted by amino acid at position (96) represented on fig. 5, or
- the one extending from the extremity constituted by amino acid at position (101) to the extremity constituted by amino acid at position (120) represented on fig. 5, or
- the one extending from the extremity constituted by amino acid at position (175) to the extremity constituted by amino acid at position (194) represented on fig. 5, or
- the one extending from the extremity constituted by amino acid at position (211) to the extremity constituted by amino acid at position (230) represented on fig. 5, or
- the one extending from the extremity constituted by amino acid at position (275) to the extremity constituted by amino acid at position (295) represented on fig. 5,

and the peptidic sequences resulting from the modification by substitution and/or by addition and/or by deletion of one or several amino acids in so far as

this modification does not alter the following properties :

the polypeptides react with rabbit polyclonal antiserum raised against the protein of 32-kDa of M. bovis BCG culture filtrate, and/or

react selectively with human sera from tuberculosis patients and particularly patients developing an evolutive tuberculosis at an early stage, and/or react with the amino acid sequence extending from the extremity constituted by amino acid at position (1), to the extremity constituted by amino acid at position (295) represented on fig. 5.

4. Recombinant polypeptide according to claim 1, containing in its polypeptidic chain, one at least of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-42) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-47) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-49) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-55) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-59) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-29) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-42) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-47) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-49) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-55) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-59) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b.

5. Recombinant polypeptide according to claim 2, containing in its polypeptidic chain, one at least of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,

- the one extending from the extremity constituted by amino acid at position (-42) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-47) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-49) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-55) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-59) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-29) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-42) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-47) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-49) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,

- the one extending from the extremity constituted by amino acid at position (-55) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-59) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b.

6. Recombinant polypeptide according to claim 3, containing in its polypeptidic chain, one at least of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (295) represented on fig. 5,
- the one extending from the extremity constituted by amino acid at position (-43) to the extremity constituted by amino acid at position (-1) represented on fig. 5,
- the one extending from the extremity constituted by amino acid at position (-30) to the extremity constituted by amino acid at position (295) represented on fig. 5,
- the one extending from the extremity constituted by amino acid at position (-43) to the extremity constituted by amino acid at position (295) represented on fig. 5.

7. Recombinant polypeptide according to claim 1, consisting in one of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (-59) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (-59) to the extremity

constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-55) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-55) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-49) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-49) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-47) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-47) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-42) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-42) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-29) to the extremity

constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,

- the one extending from the extremity constituted by amino acid at position (-29) to the extremity constituted by amino acid at position (-1) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (12) to the extremity constituted by amino acid at position (31) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (36) to the extremity constituted by amino acid at position (55) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (77) to the extremity constituted by amino acid at position (96) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (101) to the extremity constituted by amino acid at position (120) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (175) to the extremity constituted by amino acid at position (194) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (211) to the extremity constituted by amino acid at position (230) represented on fig. 3a and fig. 3b,
- the one extending from the extremity constituted by amino acid at position (275) to the extremity

constituted by amino acid at position (294) represented on fig. 3a and fig. 3b.

8. Recombinant polypeptide according to claim 2, consisting in one of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (-59) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-59) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-55) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-55) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-49) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-49) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-47) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-47) to the extremity

constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,

- the one extending from the extremity constituted by amino acid at position (-42) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-42) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-29) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (-29) to the extremity constituted by amino acid at position (-1) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (12) to the extremity constituted by amino acid at position (31) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (36) to the extremity constituted by amino acid at position (55) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (77) to the extremity constituted by amino acid at position (96) represented on fig. 4a and fig. 4b,
- the one extending from the extremity constituted by amino acid at position (101) to the extremity

constituted by amino acid at position (120) represented on fig. 4a and fig. 4b,

- the one extending from the extremity constituted by amino acid at position (175) to the extremity constituted by amino acid at position (194) represented on fig. 4a and fig. 4b,

- the one extending from the extremity constituted by amino acid at position (211) to the extremity constituted by amino acid at position (230) represented on fig. 4a and fig. 4b,

- the one extending from the extremity constituted by amino acid at position (275) to the extremity constituted by amino acid at position (294) represented on fig. 4a and fig. 4b.

9. Recombinant polypeptide according to claim 3, consisting in one of the following amino acid sequences:

- the one extending from the extremity constituted by amino acid at position (-43) to the extremity constituted by amino acid at position (295) represented on fig. 5,

- the one extending from the extremity constituted by amino acid at position (-43) to the extremity constituted by amino acid at position (-1) represented on fig. 5,

- the one extending from the extremity constituted by amino acid at position (-30) to the extremity constituted by amino acid at position (295) represented on fig. 5,

- the one extending from the extremity constituted by amino acid at position (-30) to the extremity constituted by amino acid at position (-1) represented on fig. 5,

- the one extending from the extremity constituted by amino acid at position (1) to the extremity constituted by amino acid at position (295) represented on fig. 5,

- the one extending from the extremity constituted by amino acid at position (12) to the extremity constituted by amino acid at position (31) represented on fig. 5,
- the one extending from the extremity constituted by amino acid at position (36) to the extremity constituted by amino acid at position (55) represented on fig. 5,
- the one extending from the extremity constituted by amino acid at position (77) to the extremity constituted by amino acid at position (96) represented on fig. 5,
- the one extending from the extremity constituted by amino acid at position (101) to the extremity constituted by amino acid at position (120) represented on fig. 5,
- the one extending from the extremity constituted by amino acid at position (175) to the extremity constituted by amino acid at position (194) represented on fig. 5,
- the one extending from the extremity constituted by amino acid at position (211) to the extremity constituted by amino acid at position (230) represented on fig. 5,
- the one extending from the extremity constituted by amino acid at position (275) to the extremity constituted by amino acid at position (295) represented on fig. 5.

10. Amino acid sequences constituted by a polypeptide according to claims 1 to 9, and a protein or an heterologous sequence with respect to said polypeptide, said protein or heterologous sequence comprising from about 1 to about 1000 amino acids.

11. Amino acid sequence according to claim 10, wherein the heterologous protein is  $\beta$ -galactosidase.

12. Nucleic acid comprising

- a nucleotide sequence coding for anyone of the polypeptides according to claims 1 to 11,
- or nucleotide sequences which hybridize with the nucleotide sequences coding for anyone of the polypeptides according to claims 1 to 11,
- or nucleotide sequences which are complementary to the nucleotide sequences coding for any of the polypeptides according to claims 1 to 11,
- the above mentioned nucleotide sequences wherein T can be replaced by U.

13. Nucleic acid according to claim 12, comprising one at least of the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (182) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (360) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1242) to the extremity constituted by nucleotide at position (1358), wherein N represents one of the five A, T, C, G or I nucleotides, represented in fig. 3a and fig. 3b,

or above said nucleotide sequences wherein T is replaced by U,

or nucleic acids which hybridize with said above mentioned nucleotide sequences or the complementary sequences thereof.

14. Nucleic acid according to claim 13, comprising one at least of the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (182) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (360) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1242) to the extremity constituted by nucleotide at position (1358), wherein N represents one of the five A, T, C, G or I nucleotides, represented in fig. 4a and fig. 4b,
- or above said nucleotide sequences wherein T is replaced by U,
- or nucleic acids which hybridize with said above mentioned nucleotide sequences or the complementary sequences thereof.

15. Nucleic acid according to claim 13, comprising one at least of the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (130) to the extremity constituted by nucleotide at position (219) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (220) to the extremity constituted by nucleotide at position (1104) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (1104) to the extremity constituted by nucleotide at position (1299),

or above said nucleotide sequences wherein T is replaced by U,  
or nucleic acids which hybridize with said above mentioned nucleotide sequences or the complementary sequences thereof.

16. Nucleic acid according to claim 13, comprising one of the following sequences:

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (194) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (212) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (218) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (272) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (360) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b.

17. Nucleic acid according to claim 14, comprising one of the following sequences:

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (194) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (212) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (218) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (272) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (183) to the extremity

constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (219) to the extremity

140

constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (360) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b.

18. Nucleic acid according to claim 15, comprising one of the following sequences:

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (129) represented in fig. 5,

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (219) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1104) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1299) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (90) to the extremity constituted by nucleotide at position (219) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (90) to the extremity constituted by nucleotide at position (1299) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (90) to the extremity constituted by nucleotide at position (1104) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (130) to the extremity constituted by nucleotide at position (1104) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (130) to the extremity constituted by nucleotide at position (1299) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (220) to the extremity constituted by nucleotide at position (1299) represented in fig. 5.

19. Nucleic acid according to claim 13, consisting in one of the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (182) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (194) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (212) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (218) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (272) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (359) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (360) to the extremity constituted by nucleotide at position (1241) represented in fig. 3a and fig. 3b,

- the one extending from the extremity constituted by nucleotide at position (360) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b,
- the one extending from the extremity constituted by nucleotide at position (1242) to the extremity constituted by nucleotide at position (1358) represented in fig. 3a and fig. 3b.

20. Nucleic acid according to claim 14, consisting in one of the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (182) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (194) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (212) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (218) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (272) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted

by nucleotide at position (1241) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (183) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (195) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (213) to the extremity

constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,

- the one extending from the extremity constituted by nucleotide at position (213) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (219) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (234) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (359) represented in fig. 4a and fig. 4b,
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity

constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,  
- the one extending from the extremity constituted by nucleotide at position (273) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,  
- the one extending from the extremity constituted by nucleotide at position (360) to the extremity constituted by nucleotide at position (1241) represented in fig. 4a and fig. 4b,  
- the one extending from the extremity constituted by nucleotide at position (360) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b,  
- the one extending from the extremity constituted by nucleotide at position (1242) to the extremity constituted by nucleotide at position (1358) represented in fig. 4a and fig. 4b.

21. Nucleic acid according to claim 15, consisting in one of the following nucleotide sequences:

- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (129) represented in fig. 5,  
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (219) represented in fig. 5,  
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1104) represented in fig. 5,  
- the one extending from the extremity constituted by nucleotide at position (1) to the extremity constituted by nucleotide at position (1299) represented in fig. 5,  
- the one extending from the extremity constituted by nucleotide at position (90) to the extremity constituted by nucleotide at position (219) represented in fig. 5,

- the one extending from the extremity constituted by nucleotide at position (90) to the extremity constituted by nucleotide at position (1104) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (90) to the extremity constituted by nucleotide at position (1299) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (130) to the extremity constituted by nucleotide at position (219) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (130) to the extremity constituted by nucleotide at position (1104) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (130) to the extremity constituted by nucleotide at position (1299) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (220) to the extremity constituted by nucleotide at position (1104) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (220) to the extremity constituted by nucleotide at position (1299) represented in fig. 5,
- the one extending from the extremity constituted by nucleotide at position (1104) to the extremity constituted by nucleotide at position (1299) represented in fig. 5.

22. Recombinant nucleic acid containing at least one of the nucleotide sequences according to claims 13 to 21, inserted in a heterologous nucleic acid.

23. DNA or RNA primer constituted by one of the following sequences:

A(i) CAGCTTGTGACAGGGTTCGTGGC  
A(ii) GGTCGTGGCGCCGTACG  
A(iii) CGTCGCGCGCCTAGTGTGG  
A(iv) CGGCGCCGGTCGGTGGCACGGCGA  
A(v) CGTCGGCGCGGCCCTAGTGTGG

B TCGCCCGCCCTGTACCTG

C GCGCTGACGCTGGCGATCTATC

D CCGCTGTTAACGTCGGGAAG

E AAGCCGTCGGATCTGGGTGGCAAC

F(i) ACGGCACTGGGTGCCACGCCAAC  
F(ii) ACGCCCAACACCGGGCCCGCCGCA  
F(iii) ACGGGCACTGGGTGCCACGCCAAC  
F(iv) ACGCCCCAACACCGGGCCCGCCCCCA

24. DNA or RNA primer set constituted by any of the nucleotide sequences A(i), A(ii), A(iii), A(iv), A(v), B, C, D, E, F(i), F(ii), F(iii) or F(iv) in association with the complement of any other nucleotide sequences chosen from A, B, C, D, E, or F, A meaning any of the sequences A(i), A(ii), A(iii), A(iv), A(v) and F any of the sequences F(i), F(ii), F(iii) and F(iv),

A(i), A(ii), A(iii), A(iv), A(v), B, C, D, E, F(i), F(ii), F(iii) and F(iv) having the meaning of claim 11, and

advantageously constituted by the following elements :

A(i)

or A(ii)

or A(iii)

and the complement of B

or A(iv)

or A(v)

A(i)  
or A(ii)  
or A(iii) and the complement of C  
or A(iv)  
or A(v)

B and the complement of C  
A(i)  
or A(ii)  
or A(iii) and the complement of F  
or A(iv)  
or A(v)

A(i)  
or A(ii)  
or A(iii) and the complement of D  
or A(iv)  
or A(v)

A(i)  
or A(ii)  
or A(iii) and the complement of E  
or A(iv)  
or A(v)

B and the complement of D  
B and the complement of E  
B and the complement of F  
C and the complement of D  
C and the complement of E  
C and the complement of F  
D and the complement of E  
D and the complement of F  
E and the complement of F.

25. Recombinant vector, particularly for cloning and/or expression, comprising a vector sequence,

notably of the type plasmid, cosmid or phage, and a recombinant nucleic acid according to anyone of claims 13 to 21, in one of the non essential sites for its replication.

26. Recombinant vector according to claim 25, containing in one of its non essential sites for its replication necessary elements to promote the expression of polypeptides according to anyone of claims 1 to 12 in a cellular host and possibly a promoter recognized by the polymerase of the cellular host, particularly an inducible promoter and possibly a signal sequence and/or an anchoring sequence.

27. Recombinant vector according to claim 26, containing the elements enabling the expression by E. coli of a nucleic acid according to anyone of claims 6 to 9 inserted in the vector, and particularly the elements enabling the expression of the gene or part thereof of  $\beta$ -galactosidase.

28. Cellular host which is transformed by a recombinant vector according to anyone of claims 25 to 27, and comprising the regulation elements enabling the expression of the nucleotide sequence coding for the polypeptide according to anyone of claims 1 to 12 in this host.

29. Cellular host according to claim 28, chosen from among bacteria such as E. coli, transformed by the vector according to claim 25, or chosen from among eukaryotic organism, transformed by the vector according to claim 25.

30. Expression product of a nucleic acid expressed by a transformed cellular host according to anyone of claims 28 or 29.

31. Antibody characterized by the fact that it is directed against a recombinant polypeptide according to anyone of claims 1 to 12.

32. Nucleotidic probes, hybridizing with anyone of the nucleic acids according to claims 13 to 21 or with their complementary sequences, and particularly the probes chosen among the following nucleotidic sequences.

Probes A(i), A(ii), A(iii) and A(iv)

A(i) CAGCTTGTGACAGGGTTCGTGGC  
A(ii) GGTCGTGGCGCCGTACCG  
A(iii) CGTCGCGCGCCTAGTGTCCGG  
A(iv) CGGCGCCGTGGTGGCACGGCGA  
A(v) CGTCGGCGCGGCCCTAGTGTCCGG

Probe B

TCGCCCCGCCCTGTACCTG

Probe C

GCGCTGACGCTGGCGATCTATC

Probe D

CCGCTGTTAACGTCGGGAAG

Probe E

AAGCCGTCGGATCTGGGTGGCAAC

Probes F(i) and F(ii)

F(i) ACGGCACTGGGTGCCACGCCAAC  
F(ii) ACGCCCAACACCGGGCCCGCCGA  
F(iii) ACGGGCAGTGGGTGCCACGCCAAC  
F(iv) ACGCCCCAACACCGGGCCCGGCCAAC

or their complementary nucleotidic sequences.

33. Process for preparing a recombinant polypeptide according to anyone of claims 1 to 12 comprising the following steps:

- the culture in an appropriate medium of a cellular host which has previously been transformed by an appropriate vector containing a nucleic acid according to anyone of claims 12 to 22, and
- the recovery of the polypeptide produced by the above said transformed cellular host from the above said culture medium.

34. Method for the in vitro diagnostic of tuberculosis in a patient liable to be infected by *Mycobacterium tuberculosis* comprising the following steps:

- the possible previous amplification of the amount of the nucleotide sequences according to anyone of claims 12 to 22, liable to be contained in a biological sample taken from said patient by means of a DNA primer set according to claim 24,
- contacting the above mentioned biological sample with a nucleotide probe according to claim 32, under conditions enabling the production of an hybridization complex formed between said probe and said nucleotide sequence,
- detecting the above said hybridization complex which has been possibly formed.

35. Method for the in vitro diagnostic of tuberculosis in a patient liable to be infected by *Mycobacterium tuberculosis* comprising

- contacting a biological sample taken from a patient with a polypeptide according to anyone of claims 1 to 11, under conditions enabling an in vitro immunological reaction between said polypeptide and the antibodies which are possibly present in the biological sample and
- the in vitro detection of the antigen/antibody complex which has been possibly formed.

36. Method for the in vitro diagnostic of tuberculosis in a patient liable to be infected by *M. tuberculosis*, comprising the following steps :

- the primer set and detection probe used being preferably chosen among the following elements:

Primer set

P1 GAGTACCTGCAGGTGCCGTGCCGTGATGGGCCG

P2 compl. GTACCACTCGAACGCCGGGGTGTGAT

Probe B

TCGCCCGCCCTGTACCTG

Primer set

P1 GAGTACCTGCAGGTGCCGTGCCGTGATGGGCCG

P3 compl. TCCCACTTGTAAGTCTGGCA

Probe B

TCGCCCGCCCTGTACCTG

Primer set

P1 GAGTACCTGCAGGTGCCGTGCCGTGATGGGCCG

P4 compl. CGGCAGCTCGCTGGTCAGGA

Probe B

TCGCCCGCCCTGTACCTG

Primer set

P1 GAGTACCTGCAGGTGCCGTGCCGTGATGGGCCG

P5 compl. GCGTCACCCATGCCAGGCCGATCAGG

Probe B

TCGCCCGCCCTGTACCTG or

Probe C

GCGCTGACGCTGGCGATCTATC

Primer set

P1 GAGTACCTGCAGGTGCCGTGCCGTGATGGGCCG

P6 compl. GCGCCCCAGTACTCCCAGCTGTGCGT

Probe B

TCGCCCGCCCTGTACCTG or

Probe C

GCGCTGACGCTGGCGATCTATC or

Probe D

CCGCTGTTAACGTCGGAAAG or

Probe E

AAGCCGTCGGATCTGGGTGGCAAC

Primer set

P2 ATCAACACCCCGGCGTTCGAGTGGTAC

P5 compl. GCGTCACCCATGCCAGGCCGATCAGG

Probe C

GCGCTGACGCTGGCGATCTATC

Primer set

P2 ATCAACACCCCGGCGTTCGAGTGGTAC

P6 compl. GCGCCCCAGTACTCCCAGCTGTGCGT

Probe C

GCGCTGACGCTGGCGATCTATC or

Probe D

CCGCTGTTAACGTCGGAAAG or

Probe E

AAGCCGTCGGATCTGGGTGGCAAC

Primer set

P3 TGCCAGACTTACAAGTGGGA

P5 compl. GCGTCACCCATGCCAGGCCGATCAGG

Probe C

GCGCTGACGCTGGCGATCTATC

Primer set

P3 TGCCAGACTTACAAGTGGGA

P6 compl. GCGCCCCAGTACTCCCAGCTGTGCGT

Probe C

GCGCTGACGCTGGCGATCTATC or

Probe D

CCGCTGTTAACGTCGGAAAG or

Probe E

AAGCCGTCGGATCTGGGTGGCAAC

Primer set

- contacting the biological sample with an appropriate antibody according to claim 31, under conditions enabling an in vitro immunological reaction between said antibody and the antigens of M. tuberculosis which are possibly present in the biological sample and
- the in vitro detection of the antigen/antibody complex which may be formed.

37. Necessary or kit for an in vitro diagnostic method of tuberculosis in a patient liable to be infected by *Mycobacterium tuberculosis* according to claim 34, comprising

- a determined amount of a nucleotide probe according to claim 32,
- advantageously the appropriate medium for creating an hybridization reaction between the sequence to be detected and the above mentioned probe,
- advantageously, reagents enabling the detection of the hybridization complex which has been formed between the nucleotide sequence and the probe during the hybridization reaction.

38. Necessary or kit for an in vitro diagnostic method of tuberculosis in a patient liable to be infected by *Mycobacterium tuberculosis* according to claim 35, comprising

- a polypeptide according to anyone of claims 1 to 11,
- reagents for making a medium appropriate for the immunological reaction to occur,
- reagents enabling to detect the antigen/antibody complex which has been produced by the immunological reaction, said reagents possibly having a label, or being liable to be recognized by a labeled reagent, more particularly in the case where the above mentioned polypeptide is not labeled.

39. Necessary or kit for an in vitro diagnostic method of tuberculosis in a patient liable to be

infected by *Mycobacterium tuberculosis* according to claim 36, comprising

- an antibody according to claim 31,
- reagents for making a medium appropriate for the immunological reaction to occur,
- reagents enabling to detect the antigen/antibody complexes which have been produced by the immunological reaction, said reagents possibly having a label or being liable to be recognized by a label reagent, more particularly in the case where the above mentioned antibody is not labeled.

40. Immunogenic composition comprising a polypeptide according to anyone of claims 1 to 11, in association with a pharmaceutically acceptable vehicle.

41. Vaccine composition comprising among other immunogenic principles anyone of the polypeptides according to claims 1 to 11 or the expression product of claim 30, possibly coupled to a natural protein or to a synthetic polypeptide having a sufficient molecular weight so that the conjugate is able to induce in vivo the production of antibodies neutralizing *Mycobacterium tuberculosis*, or induce in vivo a cellular immune response by activating *M. tuberculosis* antigen-responsive T cells.

42. Process for the enzymatical amplification of a nucleotide sequence according to claims 12 to 22, and detection of the amplified nucleotide sequence, wherein

- the amplification is achieved by PCR technique by means of a primer set and the detection of the PCR amplified product is achieved by a hybridization reaction with a detection probe constituted by an oligonucleotide sequence of at least 10 nucleotides, said sequence being located between the two PCR primers which have been used for amplifying said nucleotide sequence,

P4 TCCTGACCAGCGAGCTGCCG

P5 compl. GCGTCACCCATGCCAGGCCGATCAGG

Probe C

GCGCTGACGCTGGCGATCTATC

Primer set

P4 TCCTGACCAGCGAGCTGCCG

P6 compl. GCGCCCCAGTACTCCCAGCTGTGCGT

Probe C

GCGCTGACGCTGGCGATCTATC or

Probe D

CCGCTGTTAACGTCGGGAAG or

Probe E

AAGCCGTCGGATCTGGGTGGAAC

Primer set

P5 CCTGATCGGCCTGGCGATGGGTGACGC

P6 compl. GCGCCCCAGTACTCCCAGCTGTGCGT

Probe D

CCGCTGTTAACGTCGGGAAG or

Probe E

AAGCCGTCGGATCTGGGTGGAAC

or the primer set being preferably chosen among the primer sets according to claim 24, and the detection probe being constituted by any oligonucleotide sequence of at least 10 nucleotides, said sequence being located between the two PCR primers constituting the primer set which has been used for amplifying said nucleotide sequence.

43. A vector sequence forming part of a recombinant vector according to claim 25, said vector sequence having either the nucleic acid sequence represented in fig. 10b, or the nucleic acid sequence represented in fig. 11b, or the nucleic acid sequence represented in fig. 12b.

44. Plasmids comprising either the nucleic acid sequence of fig. 10b, or the nucleic acid sequence of fig. 11b, or the nucleic acid sequence of fig. 12b.

45. Peptides of claim 1, advantageously used to produce antibodies, particularly monoclonal antibodies and which have the following amino acid sequences :

Amino acid position (NH <sub>2</sub> -terminal)	Amino acid sequence	position (COOH-terminal)
12	QVPSPSMGRDIKVQFQSGGA	31
36	LYLLDGLRAQDDFSGWDINT	55
77	SFYSDWYQPACRKAGCQTYK	96
101	LTSELPGWILQANRHVKPTGS	120
175	KASDMWGPKEDEPAWQRNDPL	194
211	CGNGKPSDLGGNNLPAKFLE	230
275	KPDLQRHWVPRPTPGPPQGA	294
77	SFYSDWYQPACGKAGCQTYK	96
276	PDLQRALGATPNTGPAPQGA	299

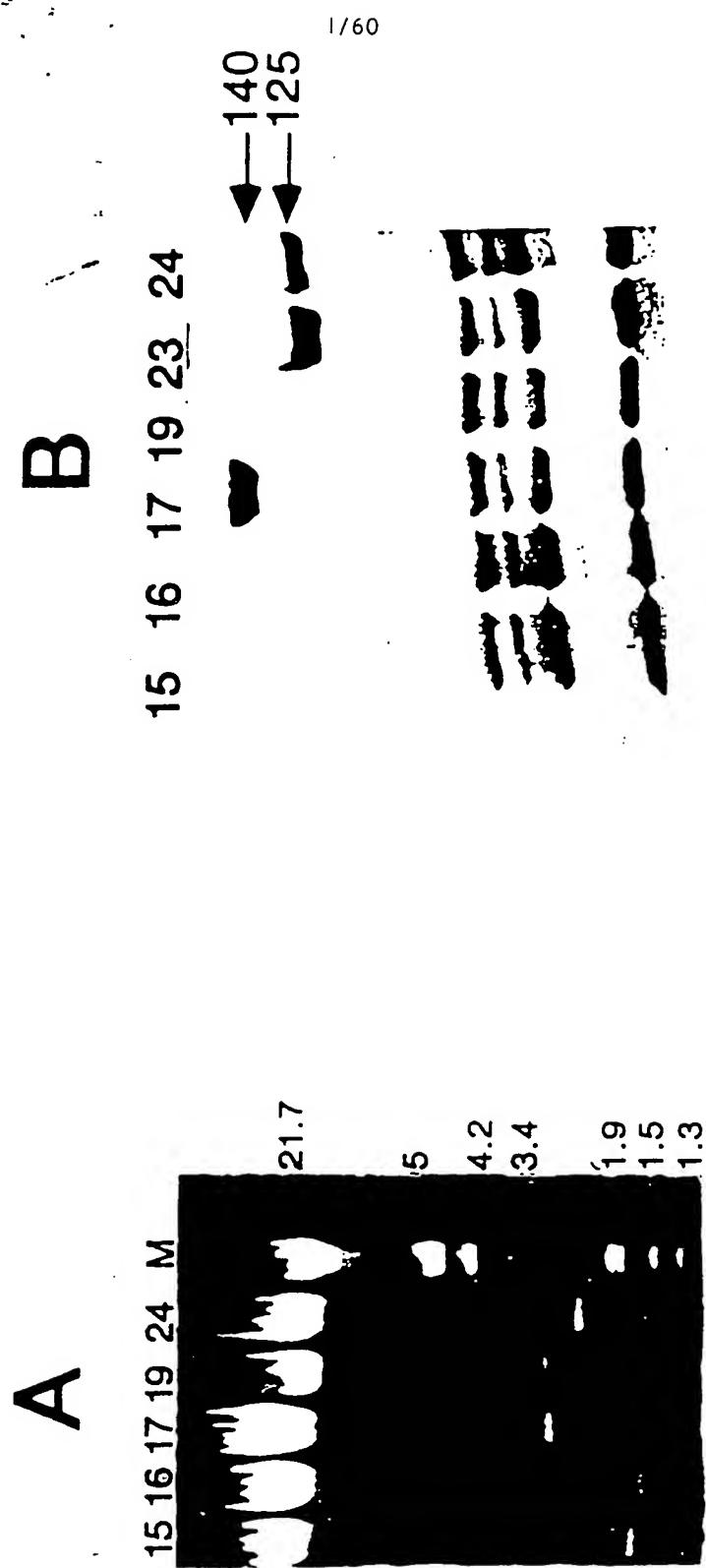
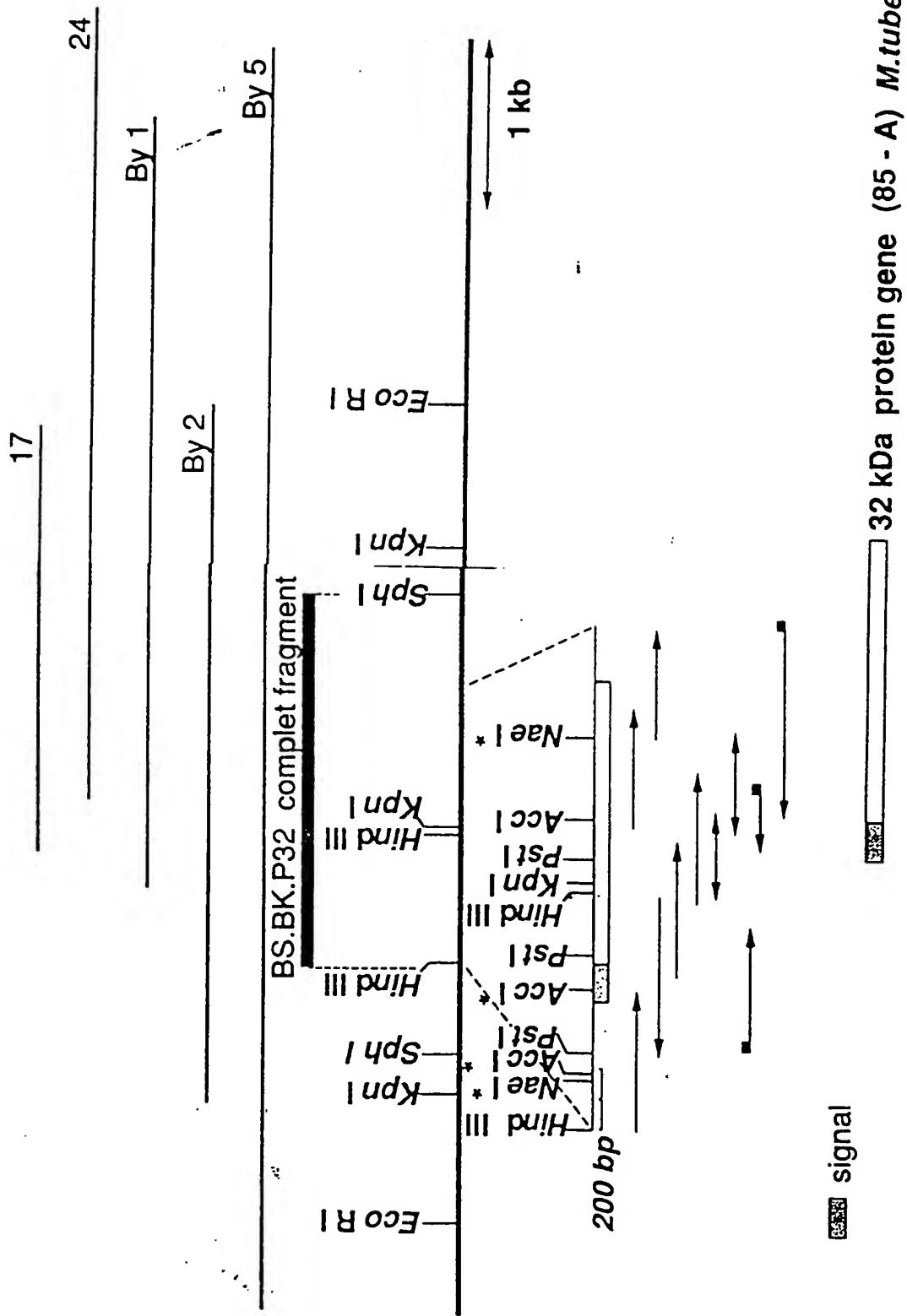


Figure 1

SUBSTITUTE SHEET

2/60  
Figure 2

CGACACATGCCAGACACTACAGGACACTGGTGGAAATGCCACCTTCAAGGGGGCTGGCTCGGGT  
 CCCGAA **TTGGC** CGTGAACGACCGCCGG **ATAA** CGGTTGGGGCTGGCTGGGGT  
 GGACGCCCAAGTTGGTGGTGGACTACAGGACACTGGCCGGCCAGGGCCCTGGCAGTCGACCT  
 ATTCAAGGATGGCCCAACATGGATGGCCTGGAGGATGAGGAGCA **AGA**  
 MET-ARG-PRO-ASN-MET-HIS-GLY-CYS-VAL-GLU-MET-ARG-MET-ARG-GLU-ALA-ARG  
 234 **AAT** G-CAG-CTT-GTT-GAC-AGG-GTT-CGT-GCC-GTC-ACG-GGT-**ATG**-TGT-CGA-CTC-GTG-GTC-  
 -42 MET-GLN-LEU-VAL-ASP-ARG-VAL-ARG-GLY-ALA-VAL-THR-GLY-MET-SER-ARG-ARG-LEU-VAL-VAL-  
 294 GGG-GCC-GTC-XCG-CXC-YTA-GTG-TCG-GGT-CTG-GTC-GCC-GCC-GTC-GGT-GGC-ACG-GCG-ACC-GCG-  
 -22 GLY-ALA-VAL-<sup>a</sup><sub>1</sub>-LEU-VAL-SER-GLY-LEU-VAL-GLY-ALA-VAL-GLY-GLY-THR-ALA-  
 354 GGG-GCA-TTT-TCC-CGG-CCG-GGC-TTG-CCG-GTG-GAG-TAC-CTG-CAG-GTG-CCG-TCG-CCG-TCG-ATG-  
 -2 GLY-ALA-phe-ser-arg-pro-gly-leu-pro-val-glu-tyr-leu-qln-val-pro-ser-pro-ser-met-  
 -1 +1  
 414 GGC-CGT-GAC-ATC-AAG)-GTC-CAA-AGT-GGT-GCC-AAC-TCG-CCC-GCC-CTG-TAC-CTG-  
 19 gly-arg-asp-ile-lys-val-gln-phe-gly-ala-asn-ser-pro-ala-leu-tyr-leu-  
 ↓ 17  
 474 CTC-GAC-GGC-CTG-CGC-GCG-CAG-GAC-GAC-TTC-AGC-GGC-TGG-GAC-AAC-AAC-CCG-GCG-TTC-  
 39 leu-asn-gly-leu-arg-ala-gln-asn-asp-asp-phe-ser-gly-trp-asp-ile-asn-thr-pro-ala-phe-  
 534 GAG-TGG-TAC-GAC-CAG-TCG-GGC-CTG-TCG-GTC-ATG-CCG-GTG-GGT-GGC-CAG-TCA-AGC-TTC-  
 59 glu-trp-tyr-asn-ser-gly-leu-ser-val-met-pro-val-gly-gly-gln-ser-ser-  
 594 TAC-TCC-GAC-TGG-TAC-CAG-CCC-GCC-TGC-7GC-AAG-GCC-GGT-TGC-CAG-(ACT-TAC-AAG-TGG-GAG-  
 79 tyr-ser-asp-trp-tyr-gln-pro-ala-cys-<sup>a</sup><sub>2</sub>-lys-ala-gly-cys-gln-thr-tyr-lys-trp-glu-

**SUBSTITUTE SHEET**

Figure 3a

654 ACC-TCG-ACC-AGC-GAG-CTG-CCG-GGG-TGG-CTG-CAG-GCC-AAC-CAC-GTC-AAG-CCC-ACC-  
 99 thr-phe-leu-thr-ser-glu-leu-pro-gly-trp-leu-gln-ala-asn-arg-his-val-lys-pro-thr-  
  
 714 GGA-AGC-GCC-GTC-GTC-GGT-CTT-TCG-ATG-GCT-GCT-TCG-GCG-CTG-ACG-CTG-GCG-ATC-TAT-  
 119 gly-ser-ala-val-gly-leu-ser-met-ala-ser-ala-ala-leu-thr-leu-ala-ile-tyr-  
  
 774 CAC-CCC-CAG-CAG-TTC-GTC-TAC-GCG-GGA-GCG-ATG-TCG-GAC-QCC-TCC-CAG-GCG-  
 139 his-pro-gln-gln-phe-val-tyr-ala-gly-ala-met-ser-gly-leu-leu-asp-pro-ser-gln-alanine-  
  
 834 ATG-GGT-CCC-ACC-CTG-ATC-GCG-CTG-GGT-GAC-GCT-GGC-GCC-TAC-AAG-GCC-TCC-GAC-  
 159 met-gly-pro-thr-ile-gly-leu-ala-met-gly-asp-ala-gly-gly-tyr-lys-ala-ser-aspartate-  
  
 894 ATG-TGG-GGC-CCG-AAG-GAG-GAC-CCG-TGG-CAG-CGC-AAC-GAC-CCG-CTG-TTG-AAC-GTC-GGG-  
 179 met-trp-gly-pro-lys-glu-asp-pro-ala-trp-gln-arg-asn-asp-pro-leu-leu-asn-val-gly-<sup>60</sup>  
  
 954 AAG-CTG-ATC-GCC-ACC-AAC-ACC-CCC-GTC-TGG-GTG-TAC-TGC-GGC-AAC-GGC-AAG-CCG-TCG-GAT-  
 199 lys-leu-ile-ala-asn-thr-arg-val-trp-val-tyr-cys-gly-asn-gly-lys-pro-ser-aspartate-  
  
 ↓ 24  
  
 1014 CTG-GGT-GGC-AAC-AAC-CTG-CCG-GCC-AAG-TTC-CTC-GAG-GGC-TTC-GTC-GGG-ACC-AGC-AAC-ATC-  
 219 leu-gly-gly-asn-asn-leu-pro-ala-lys-phe-leu-glu-gly-phe-val-arg-thr-ser-asn-ile-  
  
 1074 AAG-TTC-CAA-GAC-GCC-TAC-AAC-GCC-GGT-GCY-ZGC-CAC-AAC-GGC-GTG-TTC-GAC-CTG-GAC-  
 239 lys-phe-gln-asn-asp-ala-tyr-asn-ala-gly-gly-<sup>a2</sup>-his-asn-gly-val-phe-asp-pro-aspartate-  
  
 1134 AGC-GGT-ACG-CAC-AGC-TGG-GAG-TAC-TGG-GGC-CCG-CAG-CTC-AAC-GCT-ATG-AAG-CCC-GAC-CTG-  
 259 ser-gly-thr-his-ser-trp-glu-tyr-trp-gly-ala-gln-leu-asn-ala-met-lys-pro-asp-leu-  
  
 1242  
 1194 CAA-CG-CAC-TGG-GTG-CCG-CCA-ACA-CCG-GGC-KCL-CAG-GGC-GCC-TAGCTCCGAAACAGACA  
 279 gln-arg-<sup>a3</sup>-b<sub>3</sub>-c<sub>3</sub>-d<sub>3</sub>-e<sub>3</sub>-f<sub>3</sub>-thr-<sup>a4</sup>-gly-pro-<sup>a5</sup>-gln-gly-ala-TER  
  
 1258 CAAACATCTAGCNCNNCGGCTTGTGGNNCANATGTTTCCTAAATCCCCGTCCTAGCTCCGNNCCGTGTGGTTA  
 1338 GCTACCTGACNNCATGGGTTT 294  
 1358

## SUBSTITUTE SHEET

Figure 3b

7/60

1 ACT-GCC-GGG-CCC-AGC-GCC-TGC-AGT'-CTG-ACC-TAA-TTC-AGG-ATG-CGC-CCA-AAC-ATG-CAT'-GGA-  
 61 TGC-GTT-GAG-ATG-AGG-ATG-AGG-GAA-AGA-ATG-CAG-CTT-GTT-GAC-AGG-GTT-CGT-GGC-GCC-  
     MET-GLN-LEU-VAL-ASP-ARG-VAL-ARG-GLY-ALA-  
     (-43)

121 GTC-ACG-GGT-ATG-TCG-CGT-CGA-CTC-GTG-GTC-GGG-GCC-GCC-GTC-GCG-GCC-CTA-GTG-TCG-GGT-  
 -33 VAL-THR-GLY-MET-SER-ARG-LEU-VAL-GLY-ALA-VAL-GLY-ALA-VAL-GLY-ALA-LEU-VAL-SER-GLY-  
     +1

181 CTG-GTC-GGC-GCC-GGT-GGT-GCC-ACG-GCG-ACC-GCG-GGG-GCA-TTT-TCC-CGG-CCG-GGC-TTG-CCG-  
 -13 LEU-VAL-GLY-ALA-VAL-GLY-GLY-THR-ALA-THR-ALA-phe-ser-arg-pro-gly-leu-pro-  
     +1

241 GTG-GAG-TAC-CTG-CAG-GTG-CCG-TCG-CCG-TCG-GTC-CGT-GAC-ATC-AAG-GTC-CAA-TTC-CAA-  
 8 val-glu-tyr-leu-gln-val-pro-ser-pro-ser-pro-ser-met-gly-arg-asp-ile-lys-val-gln-phe-  
     +1

301 AGT-GGT-GGC-AAC-TCG-CCC-GCC-CTG-TAC-CTG-CTC-GAC-GGC-CTG-CGC-GCG-CAG-GAC-GAC-  
 28 ser-gly-ala-asn-ser-pro-ala-leu-tyr-leu-asp-gly-leu-arg-ala-gln-asp-asp-  
     +1

361 TTC-AGC-GGC-TGG-GAC-ATC-AAC-ACC-CCG-GCG-TTC-GAG-TGG-TAC-GAC-CAG-TCG-GGC-CTG-TCG-  
 48 phe-ser-gly-trp-asp-ile-asn-thr-pro-ala-phe-glu-trp-tyr-asp-gln-ser-gly-leu-ser-  
     +1

421 GTG-GTC-ATG-CCG-GTG-GGT-GCC-CAG-TCA-AGC-TTC-TAC-TCC-GAC-TGG-TAC-CAG-CCC-GCC-TGC-  
 68 val-val-met-pro-val-gly-gly-gln-ser-ser-phe-tyr-ser-asp-trp-tyr-asp-gln-pro-ala-cys-  
     +1

481 GGC-AAG-GCC-GGT-GTC-CAG-AC'T-TAC-AGG-TGG-GAG-ACC-CTG-ACC-AGC-GAG-CTG-CCG-GGG-  
 88 gly-lys-ala-gly-cys-gln-thr-tyr-lys-trp-glu-thr-ser-asp-trp-tyr-arg-ala-cys-  
     +1

## **SUBSTITUTE SHEET**

Figure 5

541 TGG-CTG-CAG-GCC-AAC-AGG-CAC-GTC-AAA-G-CCC-AGC-GCC-GTC-GTC-GGT-TCG-TCG-  
 108 trp-leu-gln-ala-asn-arg-his-val-lys-pro-thr-gly-ser-ala-val-gly-leu-ser-met-  
 601 GCT-GCT-TCT-TCG-GCG-CTG-ACG-CTG-GCG-ATC-TAT-CAC-CCC-CAG-CAG-CAG-TTC-GTC-TAC-GCG-GCA-  
 128 ala-ala-ser-ser-ala-leu-thr-leu-ala-ile-tyr-his-pro-gln-gln-phe-val-tyr-al-a-gly-  
 661 GCG-ATG-TCG-GGC-CTG-CTG-GAC-CCC-TCC-CAG-GCG-ATG-GGT-CCC-ACC-CTG-ATC-GGC-CTG-GCG-  
 148 ala-met-ser-gly-leu-leu-asp-pro-ser-gln-ala-met-gly-pro-thr-leu-ile-gly-leu-al-a-  
 721 ATG-GGT-GAC-GCT-GGC-TAC-AAG-GCC-TCC-GAC-ATG-TGG-GGC-CCG-AAG-GAG-GAC-CCG-GCG-  
 168 met-gly-asn-ala-gly-gly-tyr-lys-al-a-ser-asp-met-trp-gly-pro-lys-9lu-asp-pro-al-a-  
 781 TGG-CAG-CGC-AAC-GAC-CCG-CTG-TTG-AAC-GTC-GGG-AAG-CTG-ATC-GCC-AAC-AAC-ACC-CGC-GTC-  
 188 trp-gln-arg-asn-asp-pro-leu-leu-asn-val-gly-lys-leu-ile-ala-asn-asn-thr-arg-val-  
 841 TGG-GTG-TAC-TGC-GGC-AAC-CCG-TGG-GAT-CTG-GGT-GGC-AAC-AAC-CTG-CCG-GCC-AAG-  
 208 trp-val-tyr-cys-gly-asn-gly-lys-pro-ser-asp-leu-gly-gly-asn-asn-leu-pro-al-a-lys-  
 901 TTC-CTC-GAG-GGC-TTC-CTG-CGG-ACC-AGC-AAC-ATC-AAG-TTC-CAA-GAC-GCC-TAC-AAC-GCC-GGT-  
 228 phe-leu-glu-gly-phe-val-arg-thr-ser-asn-ile-lys-phe-gln-asp-ala-tyr-asn-al-a-gly-  
 961 GGC-GGC-CAC-AAC-GGC-CTG-TRC-GAC-CCG-TTC-GAC-AGC-GGT-ACG-CAC-AGC-TGG-GAG-TAC-TGG-  
 248 gly-gly-his-asn-gly-val-phe-asn-phe-pro-asp-ser-gly-thr-his-ser-trp-glu-tyr-trp-

**SUBSTITUTE SHEET**

Figure 5 (con't)

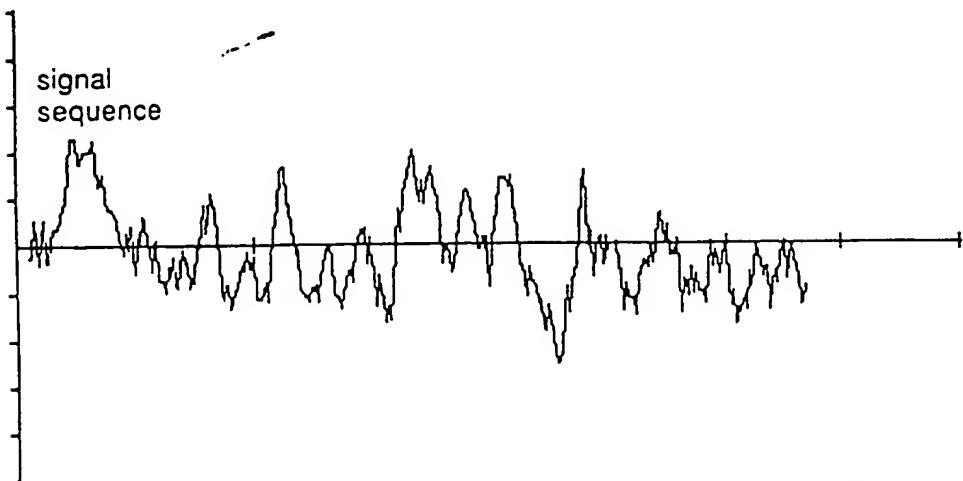
1021	GGC	C	-GCC-CAG-CTC-AAC-GCT-ATG-AAG-CCC-GAC-CTG-GCA-CGG-GCC-ACG-CCC-AAC-
268	gly	-ala	-gln-leu-asn-ala-met-lys-pro-asp-leu-gln-arg-ala-leu-gly-ala-thr-pro-asn-
			(110L)
1081	ACG	CCC	-GGG-CCC-CAG-GCC-TAG-CTC-CGA-ACA-GAC-ACA-TCT-AQC-GGC-GGT-GAC-
288	thr	-gly	-pro-ala-pro-gln-gly-ala-TER
			(295)
1141	CCT	TGT	-GGT-CGC-CGC-CGT-AGA-TGT-TTC-CTA-AAT-CCC-GTC-CCT-AGC-TCC-CGC-GGG-CCG-
1201	TGT	GGT	-TAG-CTA-CCT-GAC-GGG-CTA-GGG-GTT-GGC-GGG-GCC-GGT-TGA-CGC-CGG-GTG-CAC-ACA-
1261	GCC	TAC	-ACG-AAC-GGA-AGG-TGG-ACA-CAT-GAA-GGG-TCG-GTC

9 / 60

Figure 5 (con't)

10/60

Hydropathy

**M. tuberculosis 32 kD protein**

Hydropathy

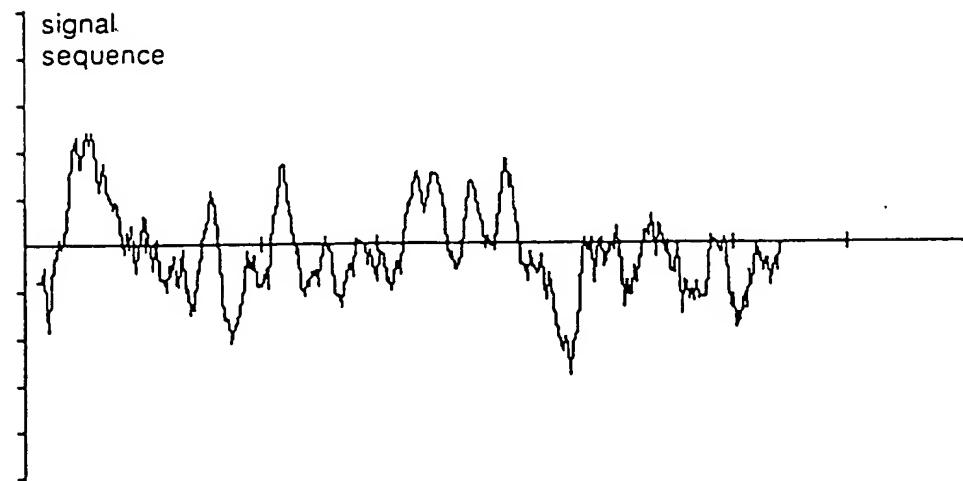
**BCG  $\alpha$ -antigen**

Fig. 6

**SUBSTITUTE SHEET**

M. tub.	<u>VDRVRGA</u> <u>VTGMSR</u> <u>RLVVGAVGAA</u> <u>LVSGLVGAVGGT</u> <u>TATAGA</u> <u>FSRPG</u> <u>GLPVEY</u> <u>LQVPS</u> <u>SPSMGR</u>	10	20	30	40	50	60
BCG	<u>MTDV</u> <u>SRKIR</u> <u>RAWGR</u> <u>RLMIGTAA</u> <u>AVVLPGLVGLAGGA</u> <u>ATAGA</u> <u>FSRPG</u> <u>GLPVEY</u> <u>LQVPS</u> <u>SPSMGR</u>	10	20	30	40	50	60
		70	80	90	100	110	120
	DIKVQFQSGGANSPALYLLDGLRAQDDFSGWDDINTPAFEWYDQSGL						
	DIKVQFQSGGNNSPAVYLLDGLRAQDDYNGWDDINTPAFEWYYQSGL						
	70	80	90	100	110	120	11/60
		130	140	150	160	170	180
	DWYQPACGKAGCQTYKWETFLTSELPGWLOANRHKPTG--SAVVG						
	DWYSPACGKAGCQTYKWETLLTSELPQWL SANRAV						
	130	140	150	160	170	180	
		190	200	210	220	230	240
	HPQQFVYAGAMSGLLDPSQAMGPTLIGLAGDAGGYKASDMWGP						
	HPQQFIYAGSLSALLDPSQGMG--LIGLAGDAGGYKAADMWGP						
	190	200	210	220	230	240	

Fig. 7

12/60

250 260 270 280 290 300  
KLIAANNTRVWVYCGNGKPSDLGGNNLPAKFQDAYNAGGGHNGNGVFD<sup>PD</sup>  
240 250 260 270 280 290  
KLVANNTRIWWVYCGNGTPNELGGANIPAAEFLENFVRSSNLKEQDAYKPA<sup>GGHNAVNFRPP</sup>  
310 320  
SGTHSWEYWGQAQLNAMKPD<sup>L</sup>QRALGA  
300 310 320  
NGTHSWEYWGQAQLNAMKGD<sup>L</sup>QSSLGA

Fig. 7 (con't)

13/60

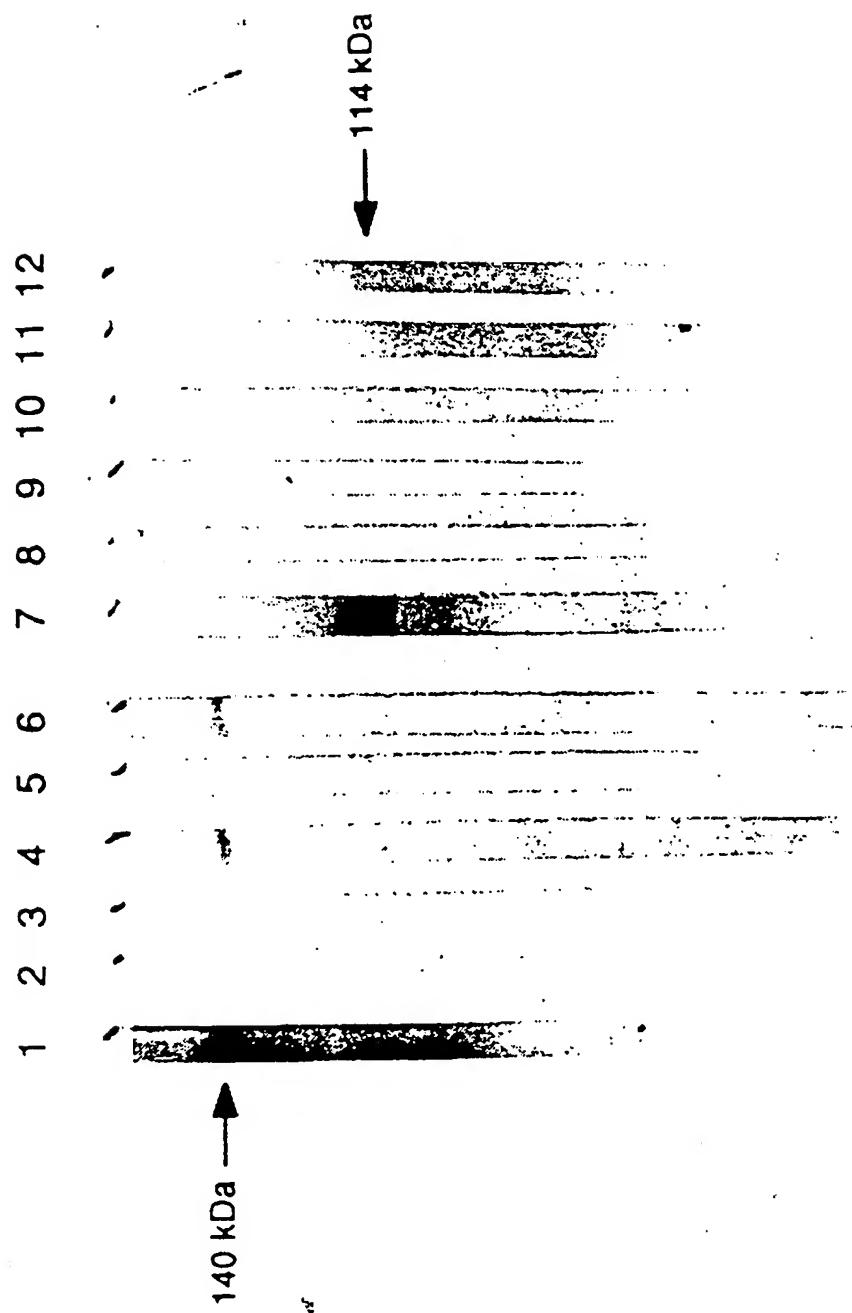
**SUBSTITUTE SHEET**

Figure 8

14/60  
Figure 9a

## PROBE REGION A

1 ATG CAGCTTGTGACAGGGTCGTGGCGCCGTACGGTATGTCGCGTCGACTC  
 1 ||| |||||||||||||||||||||||||||||||||||||||||||||||||  
 1 ATG CAGCTTGTGACAGGGTCGTGGCGCCGTACGGTATGTCGCGTCGACTC  
 1 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
 1 ATG ACAGACGTGAGCCGAAAGATTGAG CTT GGGGACGCCG ATTGA TG

- 55 GTGGTCGGGGCCGTGGCGCGCCCTAGTCGGGTCTGGTCGGCGCCGTGGTG  
 55 ||| ||||||||||| ||||| |||||||||||||||||||||||||||||  
 49 ATCGGACGGCAGCG GCTGT AGTCCTCCGGGCCTGGTGGGGCTTGCCGGCG

110 GCA CGGCGACCGCGGGGCATTTCCCGGCCGGCTTGCCTGGTG P1 GAGTACCTG  
 107 GCA CGGCGACCGCGGGGCATTTCCCGGCCGGCTTGCCTGGTG GAGTACCTG  
 101 GAG CGGCAACCGCGGGCGCTCCCGGCCGGGCTGCCGGTC GAGTACCTG

163 CAGGTGCCGTGCCGTGATGGGCCG TGACATCAAGGTCCAATTCAAAGTGGT  
 160 CAGGTGCCGTGCCGTGATGGGCCG TGACATCAAGGTCCAATTCAAAGTGGT  
 154 CAGGTGCCGTGCCGTGATGGGCCG CGACATCAAGGTTCAAGTCCAGAGCGGT

## PROBE REGION B

217 GGTGCCAAC TCGCCCGCCCTGTACCTG CTCGACGGCCTGCGCGCGAGGACGA  
 214 GGTGCCAAC TCGCCCGCCCTGTACCTG CTCGACGGCCTGCGCGCGAGGACGA  
 208 GGGAAACAAC TCACCTGCGGTTATCTG CTCGACGGCCTGCGCGCCAAAGACGA

15/60  
Figure 9b

270	CTTCAGCGGCTGGGAC	P2	ATCAACACCCCGGCCTCGAGTGGTAC	GACCAGTCGG
267	CTTCAGCGGCTGGGAC		ATCAACACCCCGGCCTCGAGTGGTAC	GACCAGTCGG
261	CTACAAACGGCTGGGAT		ATCAACACCCCGGCCTCGAGTGGTAC	TACCAAGTCGG
323	GCCTGTCGGTGGTCATGCCGGTGGGTGGCCAGTCAGCTCTACTCCGACTGGTA			
320	GCCTGTCGGTGGTCATGCCGGTGGGTGGCCAGTCAGCTCTACTCCGACTGGTA			
314	GACTGTCGATAGTCATGCCGGTGGCGGGCAGTCCAGCTCTACAGCGACTGGTA			
378	CCAGCCCGCCTGCGGCAAGGCCGGT	P3	TGCCAGACTTACAAGTGGGA	GACCT
375	CCAGCCCGCCTGCCGCAAGGCCGGT		TGCCAGACTTACAAGTGGGA	GACCT
369	CAGCCCGCCTGCCGTAAGGCTGGC		TGCCAGACTTACAAGTGGGA	AACCC
430	CTGACCAGCGAGCTGCCG	P4	TC	
427	CTGACCAGCGAGCTGCCG		TC	
421	CTGACCAGCGAGCTGCCG		TC	
484	GGAAGCGCCGTCGTCGGCTTTCGATGGCTGCTTCTTCG	PROBE REGION C	GCGCTGACGCTGGCG	
481	GGAAGCGCCGTCGTCGGCTTTCGATGGCTGCTTCTTCG		GCGCTGACGCTGGCG	
475	GGCAGCGCTGCAATCGGCTTGTGATGGCCGGCTCGTCG		GCAATGATCTTGGCC	

**SUBSTITUTE SHEET**

16/60  
Figure 9c

538 ATCTATC ACCCCCAGCAGTTCGTCTACGCCGGAGCGATGTCGGGCCTGTTGGAC  
 535 ATCTATC ACCCCCAGCAGTTCGTCTACGCCGGAGCGATGTCGGGCCTGTTGGAC  
 529 GCCTTACG ACCCCCAGCAGTTCATCTACGCCGGCTCGCTGTCGGGCCTGCTGGAC

592 CCCTCCCAGGCATGGGTCCCAC CCTGATCGGCTGGCGATGGGTGACGC TGG  
 589 CCCTCCCAGGCATGGGTCCCAC CCTGATCGGCCTGGCGATGGGTGACGC TGG  
 583 CCCTCTCAGGGGATGGG CCTGATCGGCCTCGCGATGGGTGACGC CGG

645 CGGCTACAAGGCCTCCGACATGTGGGGCCCGAAGGAGGACCCGGCGTGGCAGCGC  
 642 CGGCTACAAGGCCTCCGACATGTGGGGCCCGAAGGAGGACCCGGCGTGGCAGCGC  
 631 CGGTTACAAGGCCGCAGACATGTGGGGTCCCTCGAGTACCCGGCATGGGAGCGC

700 AACGAC CCGCTGTTAACGTCGGGAAG CTGATGCCAACACACCCGGCTCTG  
 697 AACGAC CCGCTGTTAACGTCGGGAAG CTGATGCCAACACACCCGGCTCTG  
 686 AACGAC CCTACGCCAGCAGATCCCCAAG CTGGTGGCAAACACACCCGGCTATG

753 GGTGTACTGCGGCAACGGC AAGCCGTCGGATCTGGGTGGCAAC AACCTGCCGG  
 750 GGTGTACTGCGGCAACGGC AAGCCGTCGGATCTGGGTGGCAAC AACCTGCCGG  
 739 GGTTTATTGCGGGAACGGC ACCCCGAACGAGTTGGCGGTGCC AACATACCCG

SUBSTITUTE SHEET

17/60  
 Figure 9d  
 Figure 9e

806 CCAAGTTCTCGAGGGCTTCGTGCGGACCAGCAACATCAAGTTCCAAGACGCCA  
 803 CCAAGTTCTCGAGGGCTTCGTGCGGACCAGCAACATCAAGTTCCAAGACGCCA  
 792 CCGAGTTCTGGAGAACTTCGTTCTAGCAGCAACCTGAAGTTCCAGGATGCGTA

861 CAACGCCGGTGGCGGCCACAACGGCGTGGACTTCCGGACAGCGGT P6  
 858 CAACGCCGGTGGCGGCCACAACGGCGTGGACTTCCGGACAGCGGT ACGCA  
 847 CAAGCCCGCGGGCGGGCACACGCCGTGTTCAACTTCCGCCAACGGC ACGCA

915 CAGCTGGGAGTACTGGGGCGC GCAGCTAACGCTATGAAGCCGACCTGCA AC  
 912 CAGCTGGGAGTACTGGGGCGC GCAGCTAACGCTATGAAGCCGACCTGCA AC  
 901 CAGCTGGGAGTACTGGGGCGC TCAGCTAACGCCATGAAGGGTGACCTGCAGAG

PROBE REGION F

968 GGGCACTGGGTGCCACGCCAACACCGGGCCCGGCCAGGG CGCCTAG  
 965 GGCACCTGGGTGCCACGCCAACACCGGGCC CGCCGCAGGG CGCCTAG

955 TTCGTTAGGCGCC GGCTGA

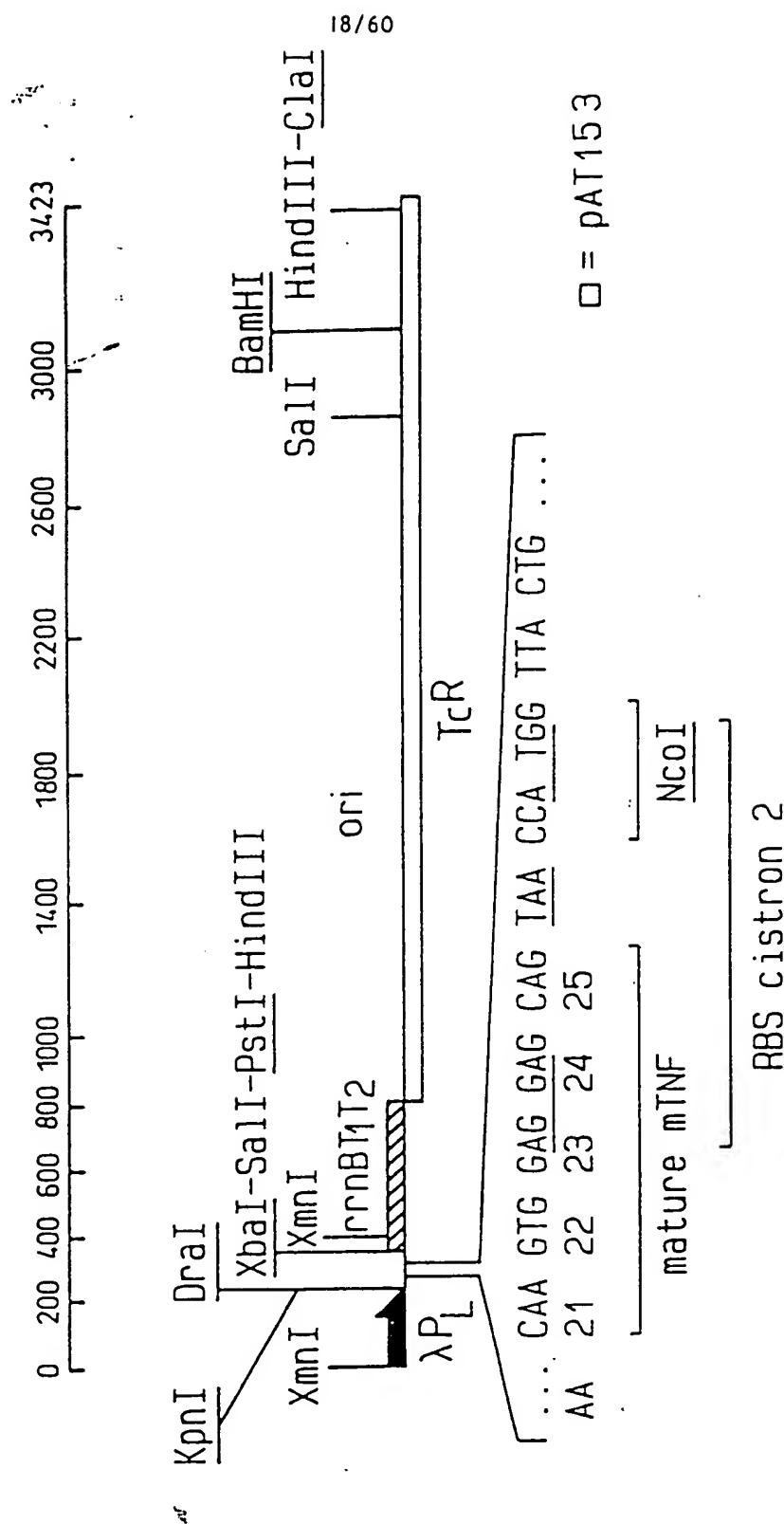


FIG. 10a

SUBSTITUTE SHEET

From: PIGRI

Fig. 10b

3	9	15	21	27	33	39	45
1	TTC	CGG	GGA	TCT	CTC	ACC	TAC
	AAG	GCC	CCT	AGA	GAG	TGG	ATG
46	TAA	ATT	CAT	ATA	AAC	ATA	CAG
	ATT	TAA	GTA	TAT	TTT	TTG	TAT
91	TTA	TCT	CTG	GCG	GTG	TTG	ACA
	AAT	AGA	GAC	CGC	CAC	AAC	TGT
136	AGC	ACA	TCA	GCA	GGA	CGC	ACT
	TCG	TGT	AGT	CGT	CCT	GGG	TGA
181	AAA	AAT	TAA	GCC	CTG	AAG	AGG
	TTT	TTA	ATT	CGG	GAC	TTC	CTG
226	TCA	TGG	TAA	GAT	CAA	GTA	GTC
	AGT	ACC	ATT	CTA	GTT	CAT	CAG
271	CCC	ACG	TCG	TAG	CAA	ACC	ACC
	GGG	TGC	AGC	ATC	GTT	TGG	TTC
316	TAC	TGG	AGA	AGG	GGG	ACC	AAC
	ATG	ACC	TCT	TCC	CCC	TGG	TTG

SUBSTITUTE SHEET

19/6

FIG. 10b (con't)

361	AGT	CTA	GAG	TCG	ACC	TGC	AGC	CCA	AGC	TTG	GCT	GTT	TTG	GGC	GAT
	TCA	GAT	CTC	AGC	TGG	ACG	TCG	GGT	TCG	AAC	CGA	CAA	AAC	CGC	CTA
406	GAG	AGA	AGA	TTT	TCA	GCC	TGA	TAC	AGA	TTA	AAT	CAG	AAC	GCA	GAA
	CTC	TCT	TCT	AAA	AGT	CGG	ACT	ATG	TCT	AAT	TTA	GTC	TTG	CGT	CTT
451	GCG	GTC	TGA	TAA	AAC	AGA	ATT	TGC	CTG	GCG	GCA	GTA	GGG	CGG	TGG
	CGC	CAG	ACT	ATT	TTG	TCT	TAA	ACG	GAC	CGC	CGT	CAT	GGC	CCC	ACC
496	TCC	CAC	CTG	ACC	CCA	TGC	CGA	ACT	CAG	AAG	TGA	AAC	GCC	GTA	GCG
	AGG	GTG	GAC	TGG	GGT	ACG	GCT	TGA	GTC	TTC	ACT	TTG	CGG	CAT	CGC
541	CCG	ATG	GTA	GTG	TGG	GGT	CTC	CCC	ATG	CGA	GAG	TAG	GGA	ACT	GCC
	GGC	TAC	CAT	CAC	ACC	CCA	GAG	GGG	TAC	GCT	CTC	ATC	CCT	TGA	CGG
586	AGG	CAT	CAA	ATA	AAA	CGA	AAG	GCT	CAG	TCG	AAA	GAC	TGG	GCC	TTT
	TCC	GTA	GTT	TAT	TTT	GCT	TTG	CGA	GTC	AGC	TTT	CTG	ACC	CGG	AAA
631	CGT	TTT	ATC	TGT	TGT	TTG	TCG	GTG	AAC	GCT	CTC	CTG	AGT	AGG	ACA
	GCA	AAA	TAG	ACA	ACA	AGC	CAC	CAC	TTG	CGA	GAG	GAC	TCA	TCC	TGT
676	AAT	CCG	CCG	GGA	GCG	GAT	TTG	AAC	GCT	CTC	CTG	CTG	AGT	AGG	ACA
	TTA	GGC	GGC	CCT	CGC	CTA	AA	TTG	CAA	CGC	TTC	GTT	GGC	CCC	GGA
721	GGG	TGG	CGG	GCA	GGA	CGC	CCG	CCA	TAA	ACT	GCC	AGG	CAT	CAA	ATT
	CCC	ACC	GCC	CGT	CCT	GGG	GGC	GGT	ATT	TGA	CGG	TCC	GTA	GTT	TAA

20/60

SUBSTITUTE SHEET

Fig. 10b (Con't)

766	AAG	CAG	AAG	GCC	ATC	CTG	ACG	GAT	GGC	CTT	TTT	GGG	T <sup>TT</sup>	CTA	C <sup>AA</sup>
	TTC	GTC	TTC	CGG	TAG	GAC	TGC	CTA	CCG	GAA	AAA	CGC	AAA	GAT	GTT
811	ACT	CTT	TTG	ATT	TTT	CTA	AAT	ACA	TTC	AAA	TAT	GTA	TCC	GCT	
	TGA	GAA	AAC	TAA	AAA	GAT	TTA	TGT	AAG	TTT	ATA	CAT	AGG	CGA	
856	CAT	GAG	ACA	ATA	ACC	CTG	ATA	ATA	GCT	TCA	ATA	ATA	AAA	GGA	TCT
	GTA	CTC	TGT	TAT	TGG	GAC	TAT	TTA	CGA	AGT	TAT	TAT	TTT	CCT	AGA
901	AGG	TGA	AGA	TCC	TTT	TTG	ATA	ATC	TCA	TGA	CCA	AAA	TCC	CTT	AAC <sup>60</sup>
	TCC	ACT	TCT	AGG	AAA	AAA	AAC	TAT	AGT	AGT	ACT	GGT	TTT	AGG	GAA
946	GTG	AGT	TTT	CGT	TCC	ACT	GAG	CGT	CAG	ACC	CCG	TAG	AAA	AGA	TCA
	CAC	TCA	AAA	GCA	AGG	TGA	CTC	GCA	GTC	TGG	GGC	ATC	TTT	TCT	AGT
991	AAG	GAT	CTT	GAG	ATC	CTT	TTT	TTC	TGC	GGC	TAA	TCT	GCT	GCT	
	TTC	CTA	GAA	GAA	CTC	TAG	GAA	AAA	AAG	ACG	CGC	ATT	AGA	CGA	CGA
1036	TGC	AAA	CAA	AAA	AAC	CAC	CGC	TAC	CAG	CGG	TGG	TTT	GTT	TGC	CGG
	ACG	TTT	GTT	TTT	TTG	GTG	GGG	ATG	GTC	GCC	ACC	AAA	CAA	ACG	GCC

Fig. 10b (con't)

1081	ATC	AAG	AGC	TAC	CAA	CTC	TTT	TTC	CGA	AGG	TAA	CTG	GCT	TCA	GCA
	TAG	TTC	TCG	ATG	GTT	GAG	AAA	AAG	GCT	TCC	ATT	GAC	CGA	AGT	CGT
1126	GAG	CGC	AGA	TAC	CAA	ATA	CTG	TCC	TTC	AGC	TGT	AGT	TAG		
	CTC	GGG	TCT	ATG	GTT	TAT	GAC	AGG	AAG	ATC	ACA	TCG	GCA	TCA	ATC
1171	GCC	ACC	ACT	TCA	AGA	ACT	CTG	TAG	CAC	CGC	CTA	CAT	ACC	TCG	CTC
	CGG	TGG	TGA	AGT	TCT	TGA	GAC	ATC	GTG	GCG	GAT	GTA	TGG	AGC	<sup>22/60</sup> GAG
1216	TGC	TAA	TCC	TGT	TAC	CAG	TGG	CTG	CTG	CCA	GTG	GCG	ATA	AGT	CGT
	ACG	ATT	AGG	ACA	ATG	GTC	ACC	GAC	GAC	GGT	CAC	CGC	TAT	TCA	GCA
1261	GTC	TTA	CCG	GGT	TGG	ACT	CAA	GAC	GAT	AGT	TAC	CGG	ATA	AGG	CGC
	CAG	AAT	GGC	CCA	ACC	TGA	GTT	CTG	CTA	TCA	ATG	GCC	TAT	TCC	CGC
1306	AGC	GGT	CGG	GCT	GAA	CGG	GGG	GTT	CGT	GCA	CAC	AGC	GCA	GCT	TGG
	TCG	CCA	GCC	CGA	CTT	GCC	CCC	CAA	GCA	CGT	GTG	TCG	GGT	CGA	ACC
1351	AGC	GAA	CGA	CCT	ACA	CCG	AAC	TGA	GAT	ACC	TAC	AGC	GTG	AGC	ATT
	TGG	CTT	GCT	GGA	TGT	GGC	TTG	ACT	CTA	TGG	ATG	TCG	CAC	TCG	TAA

Fig. 10b (con't)

1396	GAG	AAA	GCG	CCA	CGC	TTC	CCG	AAG	GAA	GAA	AGG	CGG	ACA	GGT	ATC
	CTC	TTT	CGC	GGC	GGT	GGG	AAG	GGC	TTC	CCT	TCC	GGC	TGT	CCA	TAG
1441	CGG	TAA	GCG	GCA	GGG	TCG	GAA	CAG	GAG	AGC	GCA	CGA	GGG	AGC	TTC
	GCC	ATT	CGC	CGT	CCC	AGC	CTT	GTC	CTC	TCG	CGT	GCT	CCC	TCG	<sup>23/50</sup> AAG
1486	CAG	GGG	GAA	ACG	CCT	GGT	ATC	TTT	ATA	GTC	CTG	TCG	GGT	TTC	GCC
	GTC	CCC	CTT	TGC	GGA	CCA	TAG	AAA	TAT	CAG	GAC	AGC	CCA	AAG	CGG
1531	ACC	TCT	GAC	TTG	AGC	GTC	GAT	TTT	TGT	GAT	GCT	CGT	CAG	GGG	GGC
	TGG	AGA	CTG	AAC	TGC	CAG	CTA	AAA	ACA	CTA	CGA	GCA	GTC	CCC	CCG
1576	GGA	GCC	TAT	GGA	AAA	ACG	CCA	GCA	ACG	CGG	CCT	TTT	TAC	GGT	TCC
	CCT	CGG	ATA	CCT	TTT	TGC	GGT	CGT	TGC	GCC	GGA	AAA	ATG	CCA	AGG

Fig. 10b (con't)

1621	TGG	CCT	TTT	GCT	GGC	CTT	TTG	CTC	ACA	TGT	TCT	TTC	CTG	CGT	TAT
ACC	CGA	AAA	CGA	CCG	GAA	AAC	GAG	TGT	ACA	AGA	AAG	GAC	GCA	GCA	ATA
1666	CCC	CTG	ATT	CTG	TGG	ATA	ACC	GTA	TAA	CCG	CCT	TTG	AGT	GAG	CTG
GGG	GAC	TAA	GAC	ACC	TAT	TGG	CAT	AAT	GGC	GGA	AAC	TCA	CTC	GAC	
1711	ATA	CCG	CTC	GCC	GCA	CCC	GAA	CGA	CCG	AGC	GCA	CCG	AGT	CAG	TGA
TAT	GGC	GAG	CGG	CGT	CGG	CTT	GCT	GGC	TGG	CGT	CGC	TCA	GTC	ACT <sup>24/60</sup>	
1756	GGG	AGG	AAG	CGG	AAG	AGC	GCT	GAC	TTC	CGC	GTT	TCC	AGA	CTT	TGA
CGC	TCC	TTC	GCC	TTC	TCG	TCG	CGA	CTG	AAG	GGC	CAA	AGG	TCT	GAA	ATG
1801	GAA	ACA	CGG	AAA	CCG	AGG	ACC	ATT	CAT	GTT	GTT	GCT	CAG	GTC	TAC
CTT	TGT	GGC	TTT	GGC	TTG	TCG	TGG	TAA	GTA	CAA	CAA	CGA	GTC	CAG	CGT
1846	GAC	GTT	TTG	CAG	CAG	TCG	CTT	CAC	GTT	CGC	TCG	CGT	ATC	GTC	GCA
CTG	CAA	AAC	GTC	GTC	GTC	AGC	GAA	GTG	CAA	GGG	AGC	GCA	TAG	CCA	
1891	GAT	TCA	TTC	TGC	TAA	CCA	GTA	AGG	CAA	CCC	CGC	CAG	CCT	AGC	CGG
CTA	AGT	AAG	ACG	ATT	GGT	CAT	TCC	GTT	GGG	GGG	GTC	GGA	TCG	GCC	

Fig. 10b (con't)

1936	GTC	CTC	AAC	GAC	AGG	AGC	ACG	ATC	ATG	CGC	ACC	CGT	GGC	CAG	GAC
	CAG	GAG	TRG	CTG	TCC	TCG	TCG	TGC	TAC	GCG	TGG	GCA	CCG	GTC	CTG
1981	CCA	ACG	CTG	CCC	GAG	ATG	CGC	CGC	GTG	CGG	CTG	CTG	GAG	ATG	GGC
	GGT	TGC	GAC	GGG	CTC	TAC	GGG	GGG	CAC	GCC	GAC	GAC	GAC	CTC	TAC
2026	GAC	GGC	ATG	GAT	ATG	TTC	TGC	CMA	GGG	TTC	GTT	TGC	GCA	TTC	ACA
	CTG	CGC	TAC	CTA	CTA	AAG	ACG	GTT	CCC	AAC	CAA	ACG	CGT	AAG	TGT
2071	GTT	CTC	CGC	AAG	AAT	TGA	TTG	GCT	CCA	ATT	CTT	GGA	GTG	GTG	AAT <sup>25</sup> <sub>60</sub>
	CAA	GAG	GGG	TTC	TTA	ACT	AAC	CGA	GGT	TAA	GAA	CCT	CAC	CAC	TTA
2116	CCG	TTA	GCC	AGG	TGC	CGC	CGG	CTT	CCA	TTC	AGG	TCC	AGG	TGG	CCC
	GGC	AAT	CGC	TCC	ACG	GGG	GGC	GAA	GGT	AAG	TCC	AGC	TCC	ACC	GGG
2161	GCC	TCC	ATG	CAC	CGC	GAC	GCA	ACG	CGG	GGA	GGC	AGA	CAA	GGT	ATA
	CCG	AGG	TAC	GTG	GGG	CTG	CGT	TGC	GCC	CCT	CCG	TCT	GTT	CCA	TAT
2206	GGG	CGG	CGC	CTA	CAA	TCC	ATG	CCA	ACC	CGT	TCC	ATG	TGC	TCG	CCG
	CCC	GCC	GGG	GAT	GTT	AGG	TAC	GGT	TGG	GCA	AGG	TAC	ACG	AGC	GGC

2251	AGG	CGG	CAT	AAA	TCC	CCG	TGA	CGA	TCA	GCG	GTC	CAG	TGA	TCG	NAG
	TCC	GCC	GTA	TTT	AGC	GGC	ACT	GCT	AGT	CGT	CAG	CGC	GTC	ACT	AGC
2296	TTA	GCC	TGG	TAA	GAG	CCG	CGA	GCG	ATC	CTT	GAA	GCT	GTC	CCT	GAT
	AAT	CCG	ACC	ATT	CTC	GGC	GCT	CGC	TAG	GAA	CTT	CGA	CAG	GGA	CTA
2341	GGT	CGT	CAT	CTA	CCT	GCC	TGG	ACA	GCA	TGG	CCT	GCA	ACG	CGG	GCA
	CCA	GCA	GTA	GAT	GGA	CGG	ACC	TGT	CGT	ACC	GGA	CGT	TGC	GCC	CGT
2386	TCC	CGA	TGC	CGC	CGG	AAA	CGA	GAA	TCA	TAA	TGG	GGA	AGG	CGA	
	AGG	GCT	ACG	GCG	GGC	GGC	TTC	GCT	CTT	AGT	ATT	ACC	CCT	TCC	GGT
2431	TCC	AGC	CTC	GCG	TGG	CGA	ACG	CCA	GCA	AGA	CGT	AGC	CCA	GCG	CGT
	AGG	TCG	GAG	CGC	AGC	GCT	TGC	GGT	CGT	TCT	GCA	TGG	GGT	CCG	GCA

Fig. 10b (con't)

2476	CGG	CCG	CCA	TGC	CGG	CGA	TAA	TGG	CCT	GCT	TCT	CGC	CGA	AAC	GT <sub>T</sub>
	GCC	GGC	GGT	ACG	GCC	GCT	ATT	ACC	GGA	CGA	AGA	GCG	GCT	TAT	CGA
2521	TGG	TGG	CGG	GAC	CAG	TGA	CGA	AGG	CTT	GAG	CGA	GGG	CGT	GCA	AGA
	ACC	ACC	GCC	CTG	GTC	ACT	GCT	TCC	GAA	CTC	GCT	CCC	GCA	CGT	TCT
2566	TTC	CGA	ATA	CCG	CAA	GGC	ACA	GCC	CGA	TCA	TCG	TGG	CGG	TCC	AGC
	AAG	GCT	TAT	GGC	GTT	CGC	TGT	CCG	GCT	AGT	AGC	AGC	GGG	AGG	TCG
2611	GAA	AGC	GGT	CCT	CGC	CGA	AAA	TGA	CCC	AGA	GCG	CTG	CCG	GCA	CCT <sup>27</sup> / <sub>60</sub>
	CTT	TCG	CCA	GGA	GGG	GCT	TTT	ACT	GGG	TCT	CGC	GAC	GGC	CGT	GGG
2656	GTC	CTA	CGA	GTT	GCA	TGA	TAA	AGA	CAG	TCA	TAA	GTG	CGG	CGA	
	CAG	GAT	GCT	CAA	CGT	ACT	ATT	TCT	TCT	GTC	AGT	ATT	CAC	GCC	GCT
2701	CGA	TAG	TCA	TGC	CCC	GCG	CCC	ACC	GGA	AGG	TGA	CTG	GGT	TGA	
	GCT	ATC	AGT	ACG	GGG	CGC	GGG	TGG	CCT	TCC	TCG	ACT	GAC	CCA	ACT
2746	AGG	CTC	TCA	AGG	GCA	TGG	GTC	GAC	GCT	CTC	CCT	TAT	GCG	ACT	CCT
	TCC	GAG	AGT	TCC	CGT	AGC	CAG	CTG	CGA	GAG	GGA	ATA	CGC	TGA	CGA

SUBSTITUTE SHEET

Fig. 10b (con't)

2791	GCA	TAA	GGA	AGC	CCG	CCA	GTA	GTA	GGT	GGC	GGC	TGA	GCA	CAG
	CGT	ATT	CCT	TCG	TCG	GGT	CAT	CAT	CCA	ACT	CCG	GCA	ACT	CGT
2836	CCG	CCG	CAA	GGA	ATG	GTG	CAT	GCA	AGG	AGA	TGG	CCG	CCA	ACA
	GGC	GGC	GTC	CCT	TAC	CAC	GTA	CGT	TCC	TCT	ACC	GCG	GGT	TGT
2881	CCC	CGG	CCA	CGG	GGC	CTG	CCA	CCA	TAC	CCA	CGC	CGA	ACA	MAG
	GGG	GCC	GGT	GCC	CCG	GAC	GGT	GGT	ATG	GGT	GCG	GCT	TTG	TTC
2926	TCA	TGA	GCC	CGA	AGT	GGC	GAG	CCC	GAT	CTT	CCC	CAT	CGG	TGT <sup>28/60</sup>
	AGT	ACT	CGG	GCT	TCA	CCG	CTC	GGG	CTA	GAA	GGG	GTA	GCC	ACT
2971	CGG	CGA	TAT	AGG	CGC	CAG	CAG	CAG	CAC	CTG	TGG	CCG	TGA	TGC
	GCC	GCT	ATA	TCC	GGG	GTC	GTC	GTC	GAC	GAC	ACC	GGG	CCC	ACT
3016	CGG	CCA	CGA	TGC	GTC	CGG	CGT	AGA	GGA	TCC	ACA	GGA	CGG	TGG
	GCC	GGT	GCT	ACG	CAG	GCC	GCA	TCT	CCT	AGG	TGT	CCT	CCC	CAC
3061	TCG	CCA	TGA	TCG	CGT	AGT	CGA	TAG	TGG	CTC	CAA	GTA	GCG	MAG
	AGC	GGT	ACT	AGC	GCA	TCA	GCT	ATC	ACC	GAG	GTG	CAT	CGC	TTC

## FIG. 10b (con't)

3106	GCA	GGA	CTG	GGC	GGC	GGC	GGC	CGA						
	CGT	CCT	GAC	CCG	CCG	CCG	CCG	CGT						
3151	CGG	GTG	CGC	ATA	GAA	ATT	GCA	TCA	ACG	CAT	ATA	GCG	CTA	GCA
	GCC	CAC	GCG	TAT	CTT	TAA	CGT	AGT	TGC	GTA	TAT	CGC	GAT	CGT
3196	CGC	CAT	AGT	GAC	TGG	CGA	TGC	TGT	CGG	ATG	CGA	TAT	CCC	GCA
	GGG	GTA	TCA	CTG	ACC	GCT	ACG	ACA	GCC	TTA	CCT	GCT	ATA	GGG
3241	AGA	GGC	CCG	GCA	GTA	CCG	GCA	TAA	CCA	AGC	CTA	TGC	CTA	CAG
	TCT	CCG	GGC	CGT	CAT	GGC	CGT	ATT	GGT	TCC	GAT	ACG	GAT	GTC
3286	CCA	GGG	TGA	CGG	TGC	CGA	CGA	TGA	CGA	CCG	CAT	TGT	TAG	ATT
	GGT	CCC	ACT	GCC	ACG	GCT	CCT	ACT	GCT	ACT	CGC	GTC	ACA	ATC

SUBSTITUTE SHEET

Fig. 10b (con't)

3331	TCA	TAC	ACG	GTG	CCT	GAC	TGC	GTT	AGC	AAT	TTA	ACT	GTG	ATA	ANC
	AGT	ATG	TGC	CAC	GGA	CTG	ACG	CAA	TCG	TCA	AAT	TGA	CAC	TAT	TTG
3376	TAC	CGC	ATT	AAA	GCT	TAT	CGA	TAA	GCT	GTC	AAA	CAT	GAG	ATT	
	ATG	GCG	TAA	TTT	CGA	ATA	GCT	ACT	ATT	CGA	CAG	TTT	GTA	CTC	TTA
3421	TAA														
		ATT													

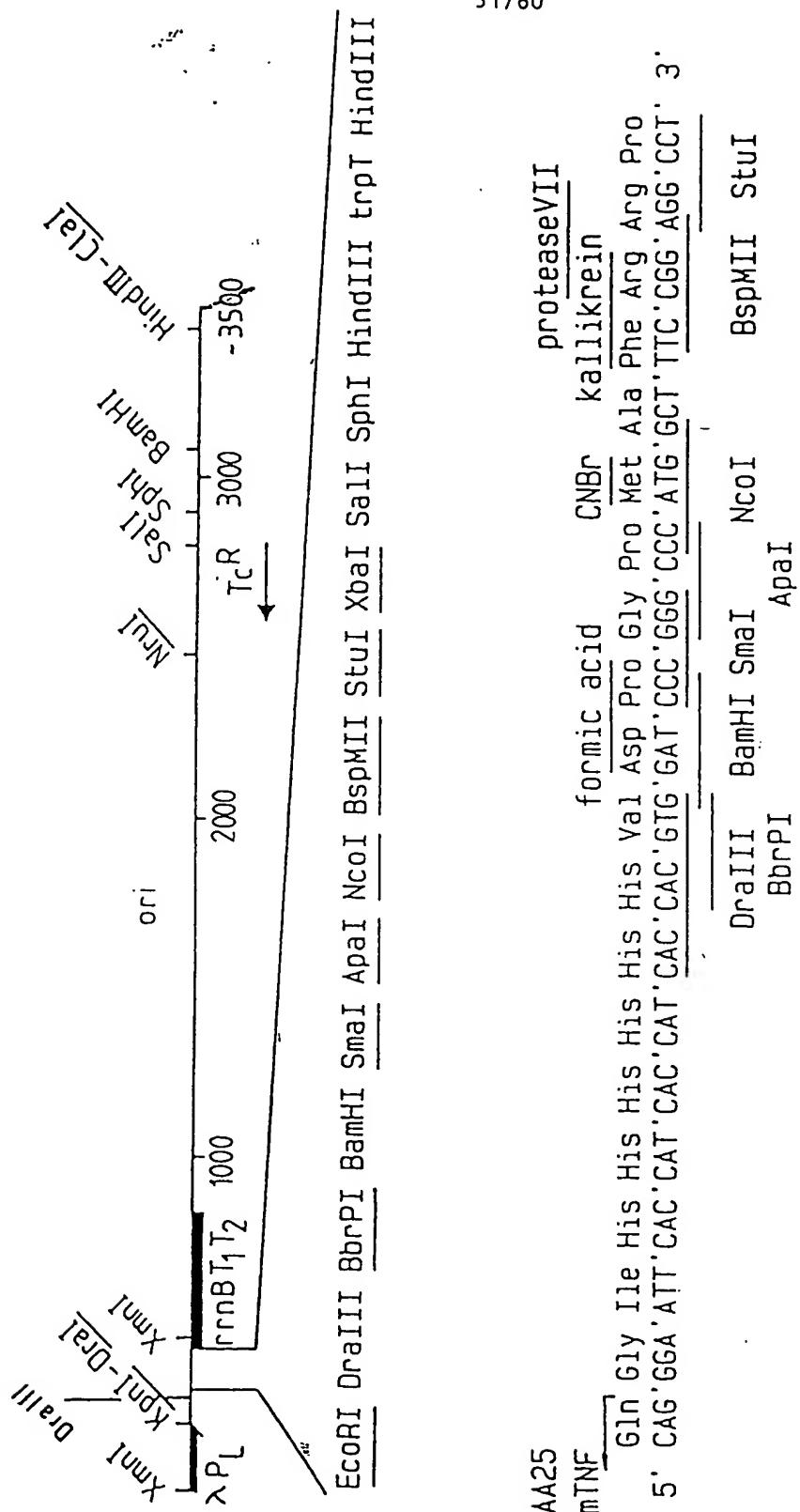
30/60

Total number of bases is: 3423.  
DNA sequence composition: 839 A; 915 C; 967 G; 702 T;

Sequence name: NIPS0060.

**SUBSTITUTE SHEET**

31/60



SUBSTITUTE SHEET

fig.11a

From: pmTNF MPH

Fig. 11b

3	9	15	21	27	33	39	45
1	AAT TCC GGG GAT CTC TCA CCT ACC AAA TGC CCC CCT GCA CGA CGT TTT						
46	AAT AAA TTC ATA TAA AAA ACA TAC AGA TAA CCA TCT GCG GTG ATA						
91	AAT TAT CTC TGG CGG TGT TGA CAT AAA TAC CAC TGG CGG TGA TAC						
136	TGA CAT CAG CAG GAC GCA CTG ACC ACC ATG GTG ACC GCC ACT ATG <sup>32/60</sup>						
181	TTA AAA ATT AAG CCC TGA AGA AGG GCA GGG GAA CCT GGT CCT CCA ATT						
226	AAT CAT GGT AAG ATC AAG TAG TCA AAA TTC GAG TGA CAA GCC TGT						
271	TTA CCA CGT CGT AGC AAA CCT AGT TCA CCT CGT CCT CGT CCC TTA AGT						
316	CCA TCA CCA TCA CCA CCT CGT CGA TCC CGG GCC CAT CCC TTT CGG GAG						
	GGT AGT GGT AGT GGT GCA CCT AGG GCC CGG GAA CGG CTC						

SUBSTITUTE SHEET

FIG. 11b (con't)

361	GCC	TCT	AGA	GTC	GAC	CGG	CAT	GCA	AGC	TTA	AGT	AGG	TAA	GCC	GCC
	CGG	AGA	TCT	CAG	CTG	GCC	GTC	CGT	TCG	AAT	TCA	TTC	ATT	CGG	CGG
406	AGT	TCC	GCT	GGC	GGC	ATT	TTN	NTT	GAT	GCC	CAA	GCT	TGG	CTG	TTT
	TCA	AGG	CGA	CCG	CCG	TAA	AAN	AAN	CTA	CGG	GTT	CGA	ACC	GAC	AAA
451	TGG	CGG	ATG	AGA	GAA	GAT	TTT	CAG	CCC	GAT	ACA	GAT	TAA	ATC	AGA
	ACC	GCC	TAC	TCT	CTT	CTA	AAA	GTC	GGG	CTA	TGT	CTA	ATT	TAG	TCT
496	ACG	CAG	AAG	CGG	TCT	GAT	AAA	ACA	GAA	TTT	GCC	TGG	CGG	CAG	TAG
	TGC	GTC	TTC	GCC	AGA	CTA	TTT	TGT	CTT	AAA	CGG	ACC	GCC	GTC	ATC
541	CGC	GGT	CCC	ACC	TGA	CCC	CAT	GCC	GAA	CTC	AGA	AGT	GAA	ACG	
	GGG	CCA	GGG	TGG	ACT	GGG	GTA	CGG	CTT	GAG	TCT	TCA	CTT	TGC	
586	CCG	TAG	CGC	CGA	TGG	TAG	TGT	GGG	GTC	TCC	CCA	TGC	GAG	AGT	AGG
	GGC	ATC	GCG	GCT	ACC	ATC	ACA	CCC	CAG	AGG	GGT	ACG	CTC	TCA	TCC
631	GAA	CTG	CCA	GGC	ATC	AAA	TAA	AAA	GAA	AGG	CTC	AGT	CGA	AAG	ACT
	CTT	GAC	GGT	CCG	TAG	TTT	ATT	TTG	CTT	TCC	GAG	TCA	GCT	TTC	TGA
676	GGG	CCT	TTC	GTT	TTA	TCT	GTT	GTT	TGT	CGG	TGA	ACG	CTC	TCC	TGA
	CCC	GGA	AAG	CAA	AAT	AGA	CAA	CAA	ACA	GCC	ACT	TGC	GAG	AGG	ACT
721	GTA	GGA	CAA	ATC	CGC	CGG	GAG	CGG	ATT	TGA	ACG	TTG	CGA	AGC	AAC
	CAT	CCT	GTT	TAG	GGG	GGC	CTC	GCC	TAA	ACT	TGC	AAC	GCT	TCG	TTG

**Fig. 11b (con't)**

766	GGC	CCG	GAG	GGT	GGC	GGC	CAG	GAC	GCC	CGC	CAT	AAA	C'G	CCA	GGC
	CCG	GGC	CTC	CCA	CCG	CCC	GGC	CTG	CGG	GGC	GAC	T'G	GGC	GGC	CCG
811	ATC	AAA	TAA	AGC	AGA	AGG	CCA	TCC	TGA	CGG	ATG	GCC	T'G	TTG	CGT
	TAG	T'T	AAT	TCG	TCT	TCC	GGT	AGG	ACT	GCC	TAC	CGG	AAA	AAC	GCA
856	T'IC	TAC	AAA	CTC	TTT	TGT	TTA	TTC	TAA	ATG	ATG	CAT	TCA	ATG	ATG
	AAG	ATG	TTT	GAG	AAA	ACA	AAA	ATG	AAA	AGC	ATT	TAT	GTA	AGT	TTA
901	TAT	CCG	CTC	ATG	AGA	CMA	TAA	CCC	TGA	TAA	ATG	C'TT	CMA	TAA	TAA
	ATA	GGC	GAG	TAC	TCT	GT'T	ATT	GGG	ACT	ATT	TAC	GAA	GTT	ATT	ATT
946	AAG	GAT	CTA	GGT	GAA	GAT	CCT	TTT	TGA	TAA	TCT	CAT	GAC	CMA	AT
	TTC	CTA	GAT	CCA	CTT	CTA	GGA	AAA	ACT	ATT	AGA	GTA	CTG	GTT	TTA
991	CCC	TTA	ACG	TGA	GTT	TTC	GTT	CCA	CTG	AGC	GTC	AGA	CCC	CGT	AGA
	GGC	AAT	TGC	ACT	CAA	AAG	CAA	GGT	GAC	TCT	CAG	TCT	GGG	GCA	TCT
1036	AAA	GAT	CAA	AGG	ATC	TTC	TTG	AGA	TCC	TTT	TCT	GGG	CGT	AT	AT
	TTT	CTA	GTT	TCC	TAG	AAG	AAC	TCT	AGG	AAA	AGA	CGC	GCA	TTA	TTA

**SUBSTITUTE SHEET**

Fig. 11b (con't)

1081	CTG	CTG	CTT	GCA	AAC	AAA	ACC	ACC	GCT	ACC	AGC	GGT	GGT	TTG	
	GAC	GAC	GAA	CGT	TTG	TTT	TGG	TGG	CGA	TGG	TCG	CCA	CCA	AAC	
1126	TTT	GCC	GGA	TCA	AGA	GCT	ACC	AAC	TCT	TTT	TCC	GAA	GGT	MAC	TGG
	AAA	CGG	CCT	AGT	TCT	CGA	TGG	TTG	AGA	AAA	AGG	CTT	CCA	TG	ACC
1171	CTT	CAG	CAG	AGC	GCA	GAT	ACC	AAA	TAC	TGT	CCT	TCT	AGT	GTA	GCC
	GAA	GTC	GTC	TCG	CGT	CTA	TGG	TTT	ATG	ACA	GGA	AGA	TCA	CAT	CGG
1216	GTA	GTT	AGG	CCA	CCA	CTT	CAA	GAA	CTC	TGT	AGC	ACC	GGC	TAC	ATA
	CAT	CAA	TCC	GGT	GGT	GAA	GTT	CTT	GAG	ACA	TCG	TGG	CGG	ATG	TAT
1261	CCT	CGC	TCT	GCT	AAT	CCT	GTT	ACC	AGT	GGC	TGC	TGC	CAG	TGG	CGA
	GGA	GCG	AGA	CGA	TTA	GGA	CAA	TGG	TCA	CCG	ACG	ACG	GTC	ACC	GCT
1306	TAA	GTC	GTG	TCT	TAC	CGG	GTT	GGA	CTC	AAG	ACG	ATA	GTT	ACC	CGA
	ATT	CAG	CAC	AGA	ATG	GCC	CAA	CCT	GAG	TTC	TGC	TAT	CAA	TGG	CCT
1351	TAA	GGC	GCA	GCG	GTC	GGG	CTG	AAC	GGG	GGG	TTC	GTG	CAC	ACA	GCC
	ATT	CCG	CGT	CGC	CAG	CCC	GAC	TTG	CCC	CCC	AAG	CAC	GTC	TGT	CGG

Fig. 11b (Con't)

1396	CAG	CTT	GGAA	GGG	AAC	GAC	CTA	CAC	CGA	ACT	GAG	ATA	CCT	ACA	GGG
	GTC	GAA	CCT	CGC	TTG	CTG	GAT	GTG	GCT	TGA	CTC	TAT	GGAA	TGT	CGC
1441	TGA	GCA	TTG	AGA	AAG	CGC	CAC	GCT	TCC	CGA	AGG	GAG	AAA	GGC	GGA
	ACT	CGT	AAC	TCT	TTC	GGC	GTG	CGA	AGG	GCT	TCC	CTC	TTT	CCG	CCT <sup>36/60</sup>
1486	CAG	GTA	TCC	GGT	AAG	CGG	CAG	GGT	CGG	AGC	AGG	AGA	GGG	CAC	GAG
	GTC	CAT	AGG	CCA	TTC	GCC	GTC	CCA	GCC	TG	TCC	TCT	CGC	GTG	CTC
1531	GGA	GCT	TCC	AGG	GGG	AAA	CGC	CTG	GTA	TCT	TAA	TAG	TCC	TGT	CGG
	CCT	CGA	AGG	TCC	CCC	TTT	GCG	GAC	CAT	AGA	AAT	ATC	AGG	ACA	GGC
1576	GTT	TCG	CCA	CCT	CTG	ACT	TGA	GCG	TCG	ATT	TTT	GTG	ATG	CTC	GTC
	CAA	AGC	GGT	GGA	GAC	TGA	ACT	CGC	AGC	TAA	AAA	CAC	TAC	GAG	CAG

Fig. 11b (con't)

1621	AGG	GGG	GGG	GAG	CCT	ATG	GMA	AAA	CCC	CAG	CAA	CGC	GGC	CTT	TTT
TCC	CCC	CGC	CTC	GGA	TAC	CTT	TTT	GCG	GTC	GTC	GAA	CCG	CCC	GAA	AAA
1666	ACG	GTT	CCT	GGC	CTT	TTG	CTG	GGC	TTT	TGC	TCA	CAT	GTT	CTT	TCC
TGC	CAA	GGA	CCG	GAA	AAC	GAC	CAC	CGG	AAA	ACG	AGT	GTA	CAA	GAA	AGG
1711	TGC	GTT	ATC	CCC	TGA	TTC	TGT	GGA	TAA	CCG	TAT	TAC	CGC	CTT	TGA
ACG	CAA	TAG	GGG	ACT	AAG	ACA	CCT	ATT	GGC	ATA	ATG	GGG	GAA	ACT	37/60
1756	GTG	AGC	TGA	TAC	CGC	TCG	CCG	CAG	CGG	MAC	GAC	CGA	GGG	CAG	CGA
CAC	TCC	ACT	ATG	GCG	AGC	GGC	GTC	GGC	GGC	TTG	CTG	GCT	GGC	GTC	GCT
1801	GTC	AGT	GAG	CGA	GGA	AGC	GGA	AGA	GCG	CTG	ACT	TCC	GGG	TTT	CCA
CAG	TCA	CTC	GCT	CCT	TCG	CCT	TCT	CGC	GAC	TGA	AAG	CGC	AAA	GGT	
1846	GAC	TTT	ACG	AAA	CAC	GGA	AAC	CGA	AGA	CCA	TTC	ATG	TTG	TTG	CTC
CTG	AAA	TGC	TTT	GTG	CCT	TTG	GCT	TCT	GGT	AAG	TAC	AAC	AAC	GAG	
1891	AGG	TCC	CAG	ACG	TTT	TGC	AGC	AGC	AGT	CGC	TTC	ACG	TTC	GCT	CGC
TCC	AGC	GTC	TGC	AAA	ACG	TCG	TCC	TCA	GGG	AGG	TGC	AAG	CGA	GGG	

SUBSTITUTE SHEET

Fig. 11b (Con't)

1936	GTA	TCG	GTG	ATT	CAT	TCT	GCT	AAC	CAG	TAA	GGC	AAC	CCC	GCC	AGC
	CAT	AGC	CAC	TAA	GTA	AGA	CGA	TGG	GTC	ATT	CGG	TTG	GGG	CGG	TCC
1981	CTA	GCC	GGG	TCC	TCA	ACG	ACA	GGA	GCA	TCA	TGC	GCA	CCC	GTG	
	GAT	CGG	CCC	AGG	AGT	TGC	TGT	CCT	CGT	GCT	AGT	ACG	CGT	GGG	CAC
2026	GCC	AGG	ACC	CAA	CGC	TGC	CCG	AGA	TGC	GCC	GCC	TGC	GGC	TGC	TGG
	CGG	TCC	TGG	GTT	GCG	ACG	GCG	TCT	ACG	CGG	CGG	ACG	CCG	ACG	ACC
2071	AGA	TGG	CGG	ACG	CGA	TGG	ATA	TGT	TCT	GCC	AAG	GGT	TGG	TTT	GCG <sup>60</sup>
	TCT	ACC	GCC	TGC	GCT	ACC	TAT	ACA	AGA	CGG	TTC	CCA	ACC	AAA	CGC
2116	CAT	TCA	CAG	TTC	GCA	AGA	ATT	GAT	TGG	CTC	CAA	TTC	TTG	TAG	
	GTA	AGT	GTC	AAG	AGG	CGT	TCT	TAA	CTA	ACC	GAG	GTT	AAG	AAC	CTC
2161	TGG	TGA	ATC	CGT	TAG	CGA	GGT	GCC	GCC	GGC	TTC	CAT	TCA	GGT	CGA
	ACC	ACT	TAG	GCA	ATC	GCT	CCA	CGG	CGG	CCG	AAG	GTA	AGT	CCA	GCT
2206	GGT	GGC	CCG	GCT	CCA	TGC	ACC	GCG	ACG	CAA	CGC	GGG	GAG	GCA	GAC
	CCA	CCG	GGC	CGA	GGT	ACG	TGG	CGC	TGC	GTG	CGG	CCC	CTC	CGT	CTG

Fig. 11b (Con't)

2251	AAG	GTA	TAG	GGC	GCC	TAC	AAT	CCA	TGC	CAA	CCC	GTT	CCA	TGT
	TTC	CAT	ATC	CCG	CCG	ATG	TTA	GGT	ACG	GTC	GGG	CAA	GGT	ACA
2296	GCT	CGC	CGA	GGC	GGC	ATA	AAT	GGC	CGT	GAC	GAT	CAG	CGG	TCC
	CGA	GGG	GCT	CCG	CCG	TAT	TTA	GGC	GCA	CTG	CTG	CTA	GTC	GCC
2341	GAT	CGA	AGT	TAG	GCT	GGT	AAG	GGC	GGC	GAG	CGA	TCC	TTG	AAG
	CTA	GCT	TCA	ATC	CGA	CCA	TTC	TCG	GGC	CTC	GCT	AGG	AAC	TTC
2386	TCC	CTG	ATG	GTC	GTC	ATC	TAC	CTG	CCT	GGA	CAG	CAT	GGC	CTG
	AGG	GAC	TAC	CAG	CAG	TAG	ATG	GAC	GGA	CCT	GTC	GTA	CCG	GAC
2431	CGC	GGG	CAT	CCC	GAT	CCC	GGA	AGC	GAG	AAG	AAT	CAT	AAT	GGG
	GGG	CCC	GTA	GGG	CTA	CGG	CGG	CCT	TCG	CTC	TTC	TTA	GTA	TTA

SUBSTITUTE SHEET

Fig. 11b (con't)

2476	GAA	GGC	CAT	CCA	GCC	TCG	CGT	CCC	GAA	CAG	CAG	GAC	GTA	CCC	
	CTT	CCG	GTA	GGT	CGG	AGC	GCA	CCG	C'TT	GCG	GTC	GTC	CAT	CGG	
2521	CAG	CCG	GTC	GGC	CGC	CAT	GCC	CCC	GAT	ATG	GGC	C'TG	CTT	CTC	GCC
	GTC	GCG	CAG	CCG	GGG	GTA	CGG	CCG	CTA	T'TA	CCG	GAC	GAA	GAG	CGG
2566	GAA	ACG	TTT	GGT	GGC	GGG	ATC	AGT	GAC	GAA	GGC	TTG	AGC	GAG	GGC
	CTT	TGC	AAA	CCA	CCG	CCC	TGG	TCA	CTG	CTT	CCG	AAC	TGG	CTC	CCG <sup>40/60</sup>
2611	GTG	CAA	GAT	TCC	GAA	TAC	CGC	AGG	CAG	CAG	GCC	GAT	CAT	CGT	CCG
	CAC	CTT	CTA	AGG	CTT	ATG	GCG	TTC	GCT	GTC	CGG	CTA	GTA	GCA	GGC
2656	GCT	CCA	GCG	AAA	GCG	GTC	C'TC	GCC	GAA	ATG	GAC	CCA	GAG	CGC	TGC
	CGA	GGT	CGC	TTT	CGC	CAG	GAG	CGG	CTT	T'TA	CTG	GGT	CTC	CCG	ACG
2701	CGG	CAC	CTG	TCC	TAC	GAG	TTG	CAT	GAT	AAA	GAA	GAC	AGT	CAT	ACG
	GCC	GTG	GAC	AGG	ATG	CTC	AAC	GTA	CTA	TTT	CTT	CTG	TCA	GTA	TTC
2746	TGC	GGC	GAC	GAT	AGT	CAT	GCC	CCC	CCC	CCC	GAA	GGA	GCT	GAC	
	ACG	CCG	CTG	CTA	TCA	GTA	CGG	GGC	GGC	GGC	CTT	CCT	CGA	CTG	

SUBSTITUTE SHEET

FIG. 11b (Con't)

2791	TGG	GTT	GAA	GGC	TCT	CAA	GGC	CAT	CGG	TGG	ACG	CTC	TCC	CTT	ATG
ACC	CAA	CTT	CCG	AGA	GTT	CCC	GTA	GCC	AGC	TGC	GAG	AGG	GAA	TAC	
2836	CGA	CTC	CTG	CAT	TAG	GAA	GCA	GCC	CAG	TAG	TAG	GTT	GAG	GCC	GTT
GCT	GAG	GAC	GTA	ATC	CTT	CCG	GGT	CGG	GTC	ATC	ATC	ATC	CTA	CTC	CGG
2881	GAG	CAC	CGC	CGC	MAG	GAA	TGG	TGC	ATG	CMA	GGA	GAT	GGC	GGC	GGC
CTC	GTG	GGC	GGC	GGC	TTC	CTT	ACC	ACG	TAC	GTT	CCT	CTA	CCG	CGG	CAA
2926	CAA	CAG	TCC	CCC	GGC	CAC	GGC	GCC	TGC	CAC	CAT	ACC	CAC	GCC	GAA
GTT	GTC	AGG	GGG	CCG	GTG	CCC	GGG	ACG	GTC	GTA	TGG	GTG	CGG	CGG	CTT <sup>60</sup>
2971	ACA	AGC	GCT	CAT	GAG	CCC	GAA	GTG	GCG	AGC	CCG	ATC	TTC	CCC	ATC
TGT	TCG	CGA	GTA	CTC	GGG	CTT	CAC	CGC	TCC	GGC	TGG	AAG	GGG	TAG	
3016	GGT	GAT	GTC	GCC	GAT	ATA	GGC	GCC	AGC	AAC	CGC	ACC	TGT	GGC	GCC
CCA	CTA	CAG	CCG	CTA	TAT	CCG	GGG	TCG	TTC	GGG	TGG	ACA	CCG	CGG	
3061	GGT	GAT	GCC	GGC	CAC	GAT	GGC	TCC	GGC	GTA	GAG	GAT	CCA	CAG	GAC
CCA	CTA	CGG	CCG	GTG	CTA	CCG	AGG	CCG	CAT	CTC	CTA	GGT	GTC	CTG	

SUBSTITUTE SHEET

Fig. 11b (Con't)

3106	GGG	TGT	GGT	CGC	CAT	GAT	CGC	GTA	GTC	GAT	AGT	GGC	TCC	AAG	TAG
CCC	ACA	CCA	GGC	GTC	GTA	CTA	GGC	CAT	CAG	CTA	TCA	CCC	AGG	TTC	ATC
3151	CGA	AGC	GAG	CAG	GAC	TGG	GCG	GCG	AAA	GCG	GTC	GGA	CAG	TGC	ACG <sup>42/60</sup>
GCT	TCG	CTC	GTC	CTG	ACC	CGC	CGC	CGG	TTT	CGC	CAG	CCT	GTC	ACG	
3196	TCC	GAG	AAC	GGG	TGC	GCA	TAG	AAA	TTG	CAT	CAA	CGC	ATA	TAG	CGC
AGG	CTC	CTC	TTG	CCC	ACG	CGT	ATC	TTT	AAC	GTA	GTT	GGG	TAT	ATC	GGC
3241	TAG	CAG	CAC	GGC	ATA	GTG	ACT	GGC	GAT	GCT	GTC	GGA	ATG	GAC	GAT
ATC	GTC	GTG	CGG	TAT	CAC	TGA	CGG	CTA	CGA	CAG	CCT	TAC	CTG	CTA	
3286	ATC	CCG	CAA	GAG	GGC	CGG	CAG	TAC	CGG	CAT	AAC	CAA	GCC	TAT	GCC
TAG	GGC	GTT	CTC	CGG	GCC	GTC	ATG	GCC	GTA	TTC	GTT	CGG	ATA	CGG	

**SUBSTITUTE SHEET**

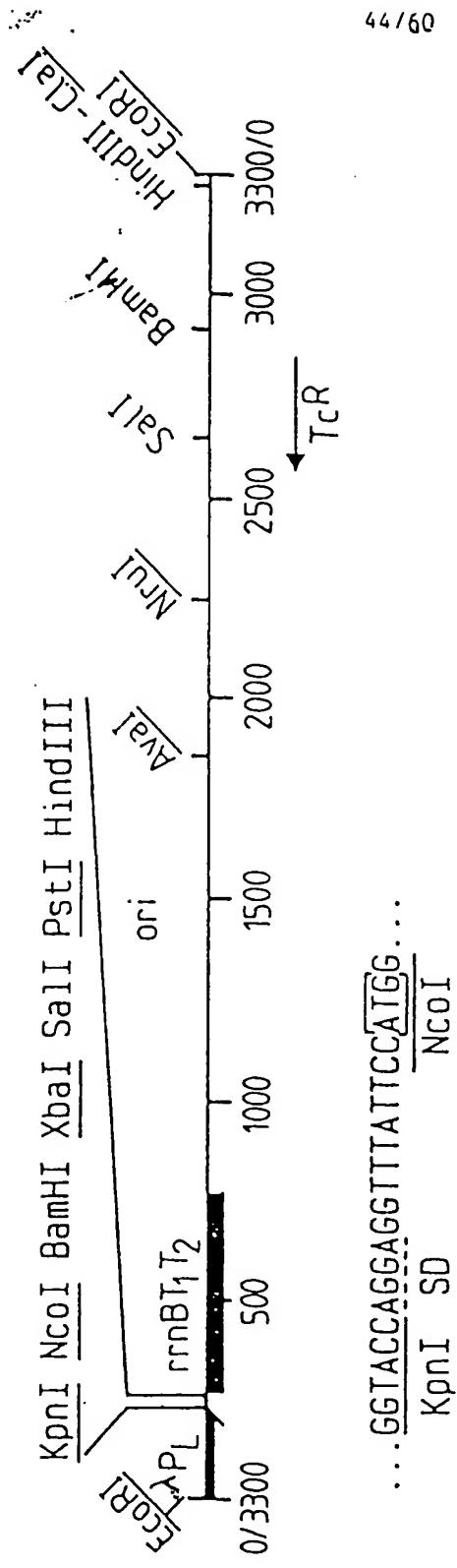
3331 TAC AGC ATC CAG GGT GAC GGC GAG GAT GAC GAG CCC ATT  
 ATG TCG TAG GTC CCA CTG CCA CGG CTC CTC CTA CTG CTC CTC GCG TAA

3376 GTT AGA TTT CAT ACA CGG TGC CTG ACT GCG TTA GCA ATT TAA CTG  
 CAA TCT AAA GTA TGT GCC ACG GAC TGA CGC ATT CGT TAA ATT GAC

3421 TGA TAA ACT ACC GCA TTA MAG CTT ATC GAT MAG C'IG TCA AAC  
 ACT ATT TGA TGG CGT AAT TTC GAA TAG CTA CTA TTC GAC AGT TTG  
 3466 ATG AGA ATT  
 TAC TCT TAA

43/60  
 Total number of bases is: 3474.  
 DNA sequence composition: 845 A; 933 C; 978 G; 716 T;  
 2 OTHER;  
 Sequence name: NPMTNFMPH.

44/60



SUBSTITUTE SHEET

fig.12a

From:	PIG2	Fig. 12b
3	9	15
1	TTC CGG GGA TCT CTC ACC TAC CAA ACA ATG CCC TGC AAA AAA	21
AAG	GCC CCT AGA GAG TGG ATG TGT TAC GGG ACG TTT TTT	27
46	TAA ATT CAT ATA AAA MAC ATA CAG ATA ACC ATC TGC GGT GAT AAA	33
ATT	TAA GTA TAT TTT TTG TAT GTC TAT TGG TAG ACG CCA CTA TTT	39
91	TTA TCT CTG GCG GTG TTG ACA TAA ATA CCA CTG GCG GTG ATA C'TG	45
AAT	AGA GAC CGC CAC AAC TGT ATT TAT GGT GAC CGC CAC TAT GAC	-
136	AGC ACA TCA GCA GGA CGC ACT GAC CAC CAT GAA GGT GAC GCT CTT	-
TCG	TGT AGT CGT CCT GCG TGA CTG GTG GTA CTT CCA CTG CGA GAA <sup>55</sup> <sub>60</sub>	-
181	AAA AAT TAA GCC CTG AAG AAG GGC AGG GGT ACC AGG AGG TTT AAA	-
TTT	TTA ATT CGG GAC TTC TTC CCG TCC CCA TGG TCC TCC AAA TTT	-
226	TAT TCC ATG GGG GGG ATC CTC TAG AGT CGA CCT GCA GCC CAA GCT	-
ATA	AGG TAC CCC CCC TAG GAG ATC TCA GCT GGA CGT CGG GTT CGA	-
271	TGG CTG TTT TGG CGG ATG AGA GAA GAT TTT CAG CCT GAT ACA GAT	-
ACC	GAC AAA ACC GCC TAC TCT CTT CTA AAA GTC GGA CTA TGT CTA	-
316	TAA ATC AGA ACG CAG MAG CGG TCT GAT AAA ACA GAA TTT GCC TGG	-
ATT	TAG TCT TGC GTC TTC GCC AGA CTA TTT TGT CTT AAA CGG ACC	-

Fig. 12b

361	CGG	CAG	TAG	CGC	GGT	CCC	ACC	TGA	CCC	CAT	GCC	GAA	CTC	AGA
	GCC	GTC	ATC	GCG	CCA	CCA	GGG	TGG	ACT	GGG	GTA	CGG	CTT	GAG
406	AGT	GAA	ACG	CCG	TAG	CGC	CGA	TGG	TAG	TGT	GGG	GTC	TCC	TCT
	TCA	CTT	TGC	GGC	ATC	GCG	GCT	ACC	ATC	ACA	CCC	CAG	AGG	GGT
451	GAG	AGT	AGG	GAA	CTG	CCA	GGC	ATC	AAA	TAA	AAA	GAA	AGG	CTC
	CTC	TCA	TCC	CTT	GAC	GGT	CCG	TAG	TTT	ATT	TTG	CTT	TCC	GAG
496	CGA	AAG	ACT	GGG	CCT	TTC	GTT	TAA	TCT	GTT	TGT	TGT	CGG	TGA
	GCT	TTC	TGA	CCC	GGA	AAG	CMA	AT	AGA	CAA	ACA	ACA	GCC	ACT
541	CTC	TCC	TGA	GGA	CMA	ATC	CGC	GGG	GAG	CGG	CGG	ATT	TGA	ACG
	GAG	AGG	ACT	CAT	CCT	GTT	TAG	GGC	GGC	CTC	GCC	TAA	ACT	TGC
586	CGA	AGC	AAC	GGC	CCG	GAG	GGT	GGC	GGG	CAG	GAC	GCC	CGC	CAT
	GCT	TCG	TTG	CCG	GGC	CTC	CCA	CCG	CCC	GTC	CTG	CGG	GCG	GAA
631	CTG	CCA	GGC	ATC	AAA	TTA	AGC	AGA	AGG	CCA	TCC	TGA	CGG	ATG
	GAC	GGT	CCG	TAG	TTT	AAT	TCG	TCT	TGT	TCC	GGT	AGG	ACT	GCC
676	TTT	TTG	CGT	TTC	TAC	AAA	CTC	TTT	TGT	TTA	TTT	TTC	TAA	ATA
	AAA	AAC	GCA	AAG	ATG	TTT	GAG	AAA	ACA	AAT	AAA	AAG	ATT	CAT
721	TCA	AAT	ATG	TAT	CCG	CTC	ATG	AGA	CAA	TAA	CCC	TGA	TAA	ATG
	AGT	TTA	TAC	ATA	GGC	GAG	TAC	TCT	GTT	ATT	GGG	ACT	ATT	CTT

Fig. 12b (con't)

766	CAA	TAA	TAA	AAG	GAT	CTA	GGT	GAA	GAT	CCT	TTT	TGA	TAA	TCT	CAT
	GTT	ATT	ATT	TTC	CTA	GAT	CCA	CTT	CTA	GGA	AAA	ACT	ATT	AGA	GTA
811	GAC	CAA	AAT	CCC	TTA	ACG	TGA	GTT	TTC	GTT	CCA	CTG	AGC	GTC	AGA
	CTG	GTT	TTA	GGG	AAT	TGC	ACT	CAA	AAG	CAA	GGA	GGT	GAC	TGG	CAG
856	CCC	CGT	AGA	AAA	GAT	CAA	AGG	ATC	TTC	TTG	AGA	TCC	TTT	TTT	TCT
	GGG	GCA	TCT	TTT	CTA	GTT	TCC	TAG	AAG	AAC	TCT	AGG	AAA	AAA	AGA <sup>47/60</sup>
901	GCG	CGT	AAT	CTG	CTG	CTT	GCA	AAC	AAA	AAA	ACC	ACC	GCT	ACC	AGC <sup>60</sup>
	GGC	GCA	TTA	GAC	GAC	GAA	CGT	TTG	TTG	TTT	TGG	TGG	CGA	TGG	TCG
946	GGT	GGT	TTG	TTT	GCC	GGA	TCA	AGA	GCT	ACC	AAC	TCT	TTT	TCC	GAA
	CCA	CCA	AAC	AAA	CGG	CCT	AGT	TCT	CGA	TGG	TTG	AGA	AAA	AGG	CTT
991	GGT	AAC	TGG	CTT	CAG	AGC	GCA	GAT	ACC	AAA	TAC	TGT	CCT	TCT	
	CCA	TTG	ACC	GAA	GTC	GTC	CGT	CTA	TGG	TTT	ATG	ACA	GGA	AGA	
1036	AGT	GTA	GCC	GTA	GTT	AGG	CCA	CCA	CTT	CAA	GAA	CTC	TGT	AGC	ACC
	TCA	CAT	CGG	CAT	CAA	TCC	GGT	GGT	GAA	GTT	CTT	GAG	ACA	TCG	TGG

Fig. 12b (Con't)

1081	GCC	TAC	ATA	CCT	CGC	TCT	GCT	ATAT	CCT	GTT	ACC	AGT	GGC	TGC	TGC
	CGG	ATG	TAT	GGA	GCG	AGA	CGA	TTA	GGA	CAA	TGG	TCA	CCG	ACG	ACG
1126	CAG	TGG	CGA	TAA	GTC	GTG	TCT	TAC	CGG	GTT	GGA	CTC	AAG	ACG	ATA
	GTC	ACC	GCT	ATT	CAG	CAC	AGA	ATG	GCC	CAA	CCT	GAG	TTC	TGC	TAT
1171	GTT	ACC	GGA	TAA	GGC	GCA	GCC	GTC	GGG	CTG	AAC	GGG	GGG	TTC	GTG
	CAA	TGG	CCT	ATT	CCG	CCT	CCC	CAG	CCC	GAC	TTG	CCC	CCC	AAG	CAC <sup>48/60</sup>
1216	CAC	ACA	GCC	CAG	CTT	GGA	GGC	MAC	GAC	CTA	CAC	CGA	ACT	GAG	ATA
	GTG	TGT	CGG	GTC	GAA	CCT	CCG	TTG	CTG	GAT	GTG	GCT	TGA	CTC	TAT
1261	CCT	ACA	GGG	TGA	GCA	TTG	AGA	AAAG	CGC	CAC	GCT	TCC	CGA	AGG	GAG
	GGA	TGT	CGC	ACT	CGT	AAC	TCT	TTC	GGG	GTG	CGA	AGG	GCT	TCC	CTC
1306	AAA	GGC	CAG	GTA	TCC	GGT	AAG	CGG	CAG	GGT	CGG	AAC	AGG	AGA	
	TTT	CCG	CCT	GTC	CAT	AGG	CCA	TTC	GCC	GTC	CCA	GCC	TTG	TCC	TCT
1351	GCG	CAC	GAG	GGA	GCT	TCC	ACG	GGG	AAA	CGC	CTG	GTA	TCT	TTA	TAG
	CGC	GTG	CTC	CCT	CGA	AGG	TCC	CCC	TTT	GCG	GAC	CAT	AGA	AAT	ATC

SUBSTITUTE SHEET

FIG. 12b (Con't.)

1396	TCC	TGT	CGG	GTT	TCG	CCA	CCT	CTG	ACT	TGA	GCG	TCG	ATT	TTT	GTG	
	AGG	ACA	GCC	CAA	AGC	GGT	GGA	GAC	TGA	ACT	CGC	AGC	TAA	AAA	CAC	
1441	ATG	CTC	GTC	AGG	GGG	GCG	GAG	CCT	ATG	GAA	AAA	CGG	CAG	CAA	CGC	
	TAC	GAG	CAG	TCC	CCC	CCG	CTC	GGA	TAC	CTT	TTT	GCG	GTC	GTT	GGC	
1486	GGC	CTT	TTT	ACG	GTT	CCT	GGC	CTT	TTC	CTG	GCC	TTT	TGC	TCA	CAT <sup>49/5</sup>	
	CCG	GAA	AAA	TGC	CAA	GGG	CCG	GAA	AAA	AGC	GAC	CGG	AAA	ACG	AGT	GTA
1531	GTT	CTT	TCC	TGC	GTT	ATC	CCC	TGA	TTC	TGT	GGA	TAA	CCG	TAT	TAC	
	CAA	GAA	AGG	ACG	CAA	TAG	GGG	ACT	AAG	ACA	CCT	ATT	GGC	ATA	ATG	
1576	CGC	CTT	TGA	GTG	AGC	TGA	TAC	CGC	TCG	CCG	CAG	CCG	AAC	GAC	CGA	
	GGG	GAA	ACT	CAC	TCG	ACT	ATG	GGC	AGC	GGC	GTC	GGC	TTG	CTG	GCT	

Fig. 12b (con't)

1621	GGG	CAG	CGA	GTC	AGT	GAG	CGA	GGA	AGA	GCG	CTG	ACT	TCC
	CGC	GTC	GCT	CAG	TCA	CTC	GCT	CCT	TCG	CCT	GAC	TGA	AGG
1666	GGG	TTT	CCA	GAC	TTT	ACG	AAA	CAC	GGA	AAC	CGA	AGA	CCA
	CGC	AAA	GGT	CTG	AAA	TGC	TTT	GTG	CCT	TTC	GCT	TCT	GGT
1711	TTG	TTG	CTC	AGG	TCG	CAG	ACG	TTT	TGC	AGC	AGC	AGT	CGC
	AAC	AAC	GAG	TCC	AGC	GTC	TGC	AAA	ACG	TCG	TCG	TCA	GGG
1756	TTC	GCT	CGC	GTA	TCG	GTG	ATT	CAT	TCT	GCT	AAC	CAG	TAA
	AAG	CGA	GCG	CAT	AGC	CAC	TAA	GTA	AGA	CGA	TTG	GTC	ATT
1801	CCC	GGC	AGC	CTA	GGC	GGG	TCC	TCA	ACG	ACA	GGA	GCA	CGA
	GGG	CGG	TCG	GAT	CGG	CCC	AGG	AGT	TGC	TGT	CCT	CGT	TCA
1846	GCA	CCC	GTG	GCC	AGG	ACC	CAA	CGC	TGC	CCC	AGA	TGC	GCC
	CGT	GGG	CAC	CGG	TCC	TGG	GTT	GGG	ACG	GGC	TCT	ACG	CGG
1891	GGC	TGC	TGG	AGA	TGG	CGG	ACG	CGA	TGG	ATA	TGT	TCT	GCC
	CCG	ACG	ACC	TCT	ACC	GCC	TGC	GCT	ACG	TAT	ACA	AGA	CGG

SUBSTITUTE SHEET

Fig. 12b (con't)

1936	TGG	TTT	GCG	CAT	TCA	CAG	TTC	TCC	GCA	AGA	ATT	GAT	TGG	CTC	CAA
ACC	AAA	CGC	GTA	AGT	GTC	AAG	AGG	CGT	TCT	TAA	CTA	ACC	GAG	GTT	
1981	TTC	TTG	GAG	TGG	TGA	ATC	CGT	TAG	CGA	GGT	GCC	GCC	TTC	CAT	
AAG	AAC	CTC	ACC	ACT	TAG	GCA	ATC	GCT	CCA	CGG	CGG	CCG	AAG	GTA	
2026	TCA	GGT	CGA	GGC	CCG	GCT	CCA	TGC	ACC	GGG	ACG	CAA	CGC	GGG	
AGT	CCA	GCT	CCA	CCG	GGC	CGA	GGT	ACG	TGG	CGG	TGC	GTT	GGG	CCC	
2071	GAG	GCA	GAC	AAG	GTA	TAG	GGC	GGC	GCC	TAC	AAT	CCA	TGC	CAA	CCC <sub>51/60</sub>
CTC	CGT	CTG	TTC	CAT	ATC	CCG	CCG	CCG	CGG	ATG	TTA	GGT	ACG	GTT	GGG
2116	GTT	CCA	TGT	GCT	CGC	CGA	GGC	GGC	ATA	AAT	CGC	CGT	GAC	GAT	CAG
CAA	GGT	ACA	CGA	GGC	GCT	CCG	CCG	CCG	TAT	TTA	GGG	GCA	CTG	CTA	GTC
2161	CGG	TCC	AGT	GAT	CGA	AGT	TAG	GCT	GGT	AAG	AGC	GGC	GAG	CGA	TCC
GCC	AGG	TCA	CTA	GCT	TCA	ATC	ATC	CGA	CCA	TTC	TCG	GGG	CTC	GCT	AGG
2206	TTG	AAG	CTG	TCC	CTG	ATG	GTC	GTC	ATC	TAC	CTG	CCT	GGA	CAG	CAT
AAC	TTC	GAC	AGG	GAC	TAC	CAG	CAG	CAG	ATG	GAC	GGA	CCT	GTC	GTA	

FIG. 12b (con't)

2251	GGC	CTG	CAA	CGC	GGG	CAT	CCC	GAT	GCC	GGC	AGC	GAG	AAG	AT	
	CCG	GAC	GTC	GGC	CCC	GTA	GGG	CTA	CGG	CCT	TCC	CTC	TTC	TTA	
2296	CAT	AAT	GGG	GAA	GGC	CAT	CCA	GCC	TCG	CGT	CGC	GAA	CGC	CAG	CAA
	GTA	TTA	CCC	CTT	CCG	GTA	GGT	CGG	AGC	GCA	GCG	CTT	GCG	GTC	GTT <sup>52/60</sup>
2341	GAC	GTA	GCC	CAG	CGC	GTC	GGC	CAG	CAT	GGC	GGC	GAT	AT	GGC	CTG
	CTG	CAT	CGG	GTC	GGC	CAG	CCG	GGC	CTA	GGG	CCG	CTA	TTA	CCG	GAC
2386	CTT	CTC	GCC	GAA	ACG	TTT	GGT	GGC	GGG	ACC	AGT	GAC	GAA	GGC	TTG
	GAA	GAG	CGG	CTT	TGC	AAA	CCA	CCG	CCC	TGG	TCA	CTG	CTT	CCG	AAC
2431	AGC	GAG	GGC	GTG	CAA	GAT	TCC	GAA	TAC	CCG	AAG	CGA	CAG	GCC	GAT
	TCG	CTC	CCG	CAC	GTT	CTA	AGG	CTT	ATG	GCG	TTC	GCT	GTC	CGG	CTA

Fig. 12b (con't)

2476	CAT	CGT	CGC	GCT	CCA	CCG	AAA	GCG	GTC	CTC	GCC	GAA	AAT	GAC	CCA
	GTA	GCA	GCG	CGA	GGT	CGC	TTT	CGC	CAG	GAG	CGG	CTT	TTA	CTG	GGT
2521	GAG	CGC	TGC	CGG	CAC	CTG	TCC	TAC	GAG	TTG	CAT	GAT	AAA	GAA	GAC
	CTC	GGG	ACG	GCC	GTG	GAC	AGG	ATG	CTC	AAC	GTA	CTA	TTT	CTT	CTG
2566	AGT	CAT	AAG	TGC	GGC	GAC	GAT	AGT	CAT	GCC	CCG	CCG	CCA	CCG	GAA
	TCA	GTA	TTC	ACG	CCG	CTG	CTA	CTG	CTA	GTA	CGG	GGC	GGT	GGC	CTT
2611	GGA	GCT	GAC	TGG	GTT	GAA	GGC	TCT	CAA	GGG	CAT	GGG	TCG	ACG	CTC <sup>53/60</sup>
	CCT	CGA	CTG	ACC	CAA	CTT	CCG	AGA	GTT	CCC	GTA	GCC	AGC	TGC	GAG
2656	TCC	CTT	ATG	CGA	CTC	CTG	CAT	TAG	GAA	GCA	GCC	CAG	TAG	TAG	GTT
	AGG	GAA	TAC	GCT	GAG	GAC	GTA	ATC	CTT	CGT	CGG	GTC	ATC	ATC	CAA
2701	GAG	GCC	GTG	GAG	CAC	CGC	CGC	AAG	GAA	TGG	TGG	ATG	CAA	GGA	
	CTC	CGG	CAA	CTC	GTG	GGG	GGG	GGG	TTC	CTT	ACC	ACG	TAC	GTT	CCT
2746	GAT	GGC	GCC	CAA	CAG	TCC	CCC	GGC	CAC	GGG	GGG	TGG	CAC	CAT	ACC
	CTA	CCG	CGG	GTG	GTC	AGG	GGG	CCG	GTG	CCC	GGG	ACG	GTG	GTA	TGG

Fig. 12b (con't)

2791	CAC	GGC	GAA	ACA	AGC	GCT	CAT	GAG	CCC	GAA	GTG	GGC	AGC	CCG	ATC
	GTG	CGG	CTT	TGT	TGT	TCG	CGA	GTA	CTC	GGG	CTT	CAC	CGC	TCG	GGC
2836	TTC	CCC	ATC	GGT	GAT	GTC	GGC	GAT	ATA	GGC	GCC	AGC	AAC	CGC	ACC
	AAG	GGG	TAG	CCA	CTA	CAG	CCG	CTA	TAT	CCG	CGG	TGG	TTG	GCG	TGG
2881	TGT	GGC	GCC	GGT	GAT	GCC	GGC	CAC	GAT	GCG	TCC	GGC	GTA	GAG	GAT
	ACA	CCG	CGG	CCA	CTA	CGG	CCG	GTG	CTA	CGC	AGG	CCG	CAT	CTC	CTA
2926	CCA	CAG	GAC	GGG	TGT	GGT	GGC	CAT	GAT	GGC	GTA	GTC	GAT	AGT	GGC <sup>54</sup> <sub>60</sub>
	GTC	GTC	CTG	CCC	ACA	CCA	GGC	GTC	CTA	GGC	CAT	CAG	CTA	TCA	CCG
2971	TCC	AAG	TAG	CGA	AGC	GAG	CAG	GAC	TGG	GCG	GCC	AAA	GCG	CTC	
	AGG	TTC	ATC	GCT	GCT	TCG	CTC	GTC	CTG	ACC	CGC	CGC	CGG	TTT	CCG
3016	GGA	CAG	TGC	TCC	GAG	AAC	GGG	TGC	GCA	TAG	AAA	TTG	CAT	CAA	CGC
	CCT	GTC	ACG	AGG	CTC	TTG	CCC	ACG	CGT	ATC	TTT	AAC	GTA	GTT	CCG
3061	ATA	TAG	CGC	TAG	CAG	CAC	GCC	ATA	GTG	ACT	GGC	GAT	GCT	GTC	GGA
	TAT	ATC	GGG	ATC	GTC	GTG	CGG	TAT	CAC	TGA	CCG	CTA	CGA	CAG	CCT

55/60

FIG. 12b (Con't)

3106	ATG	GAC	GAT	ATC	CCG	CAA	GAG	GCC	CGG	CAG	TAC	CGG	CAT	AAC	CAA
	TAC	CTG	CTA	TAG	GGC	GGT	CTC	CGG	GCC	GTC	ATG	GCC	GTA	TTG	GTT
3151	GCC	TAT	GCC	TAC	AGC	ATC	CAG	GGT	GAC	GGT	GCC	GAG	GAT	GAC	GAT
	CGG	ATA	CGG	ATG	TCG	TAG	GTC	CCA	CTG	CCA	CGG	CTC	CTA	CTG	CTA
3196	GAG	CGC	ATT	GTT	AGA	TTT	CAT	ACA	CGG	TGC	CTG	ACT	GCG	TTA	GCA
	CTC	GGC	TAA	CAA	TCT	AAA	GTA	TGT	GCC	ACG	GAC	TGA	CCG	AAT	CGT
3241	ATT	TAA	CTG	TGA	TAA	ACT	ACC	GCA	TTA	AAG	CTT	ATC	GAT	GAT	AAG
	TAA	ATT	GAC	ACT	ATT	TGA	TGG	CGT	AAT	TTC	GAA	TAG	CTA	CTA	TTC
3286	CTG	TCA	AAC	ATG	AGA	A									
		GAC	AGT	TTG	TAC	TCT	T								

Total number of bases is: 3301.  
 DNA sequence composition: 797 A; 887 C; 936 G; 681 T;  
 Sequence name: NIPS0039.

56/60

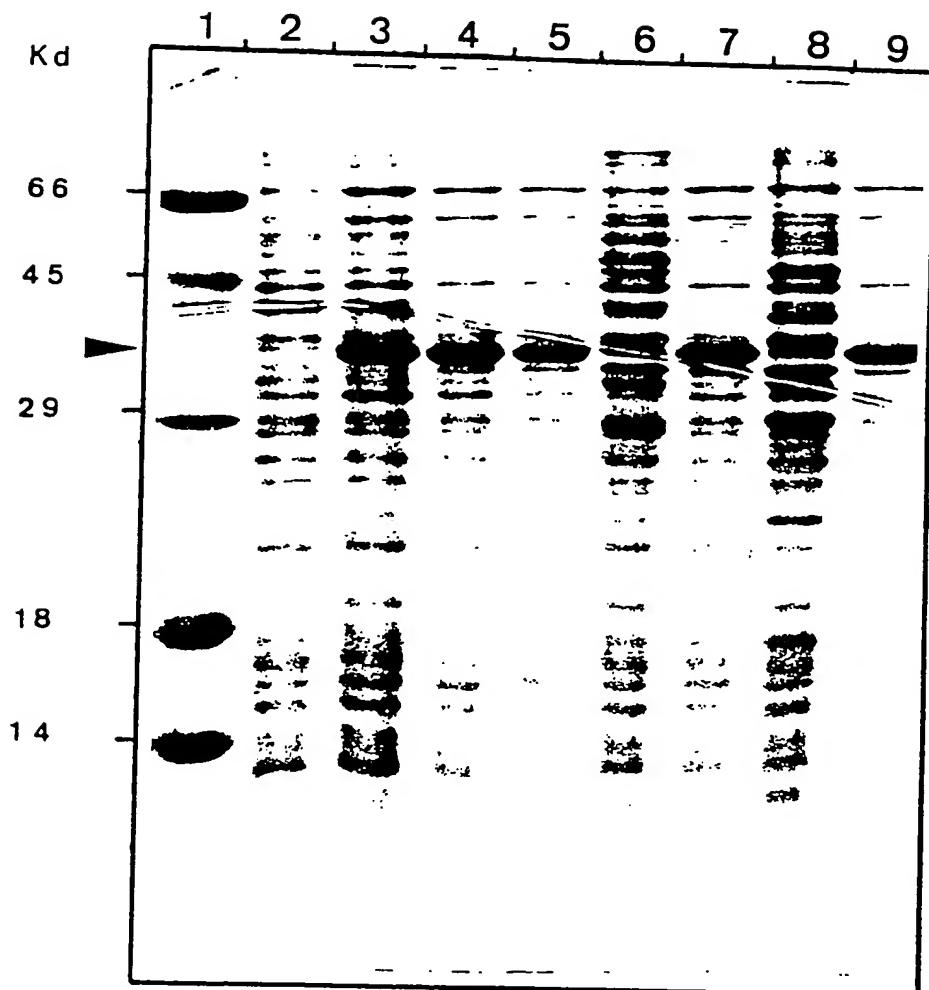


fig. 14a

SUBSTITUTE SHEET

57/60

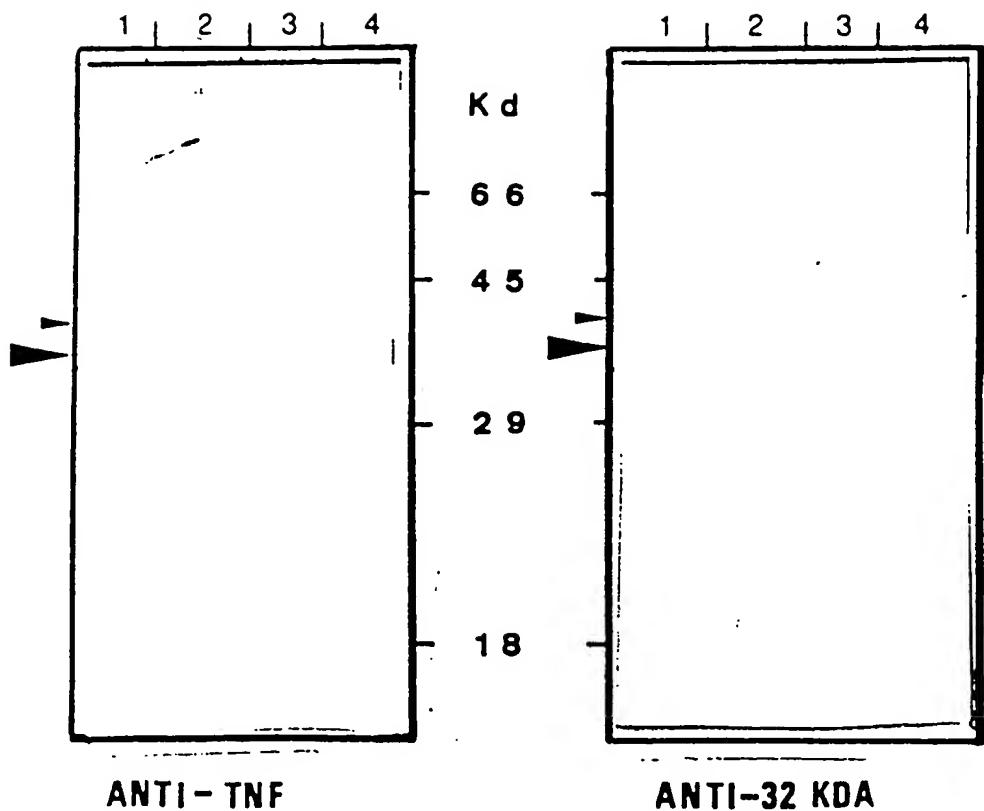
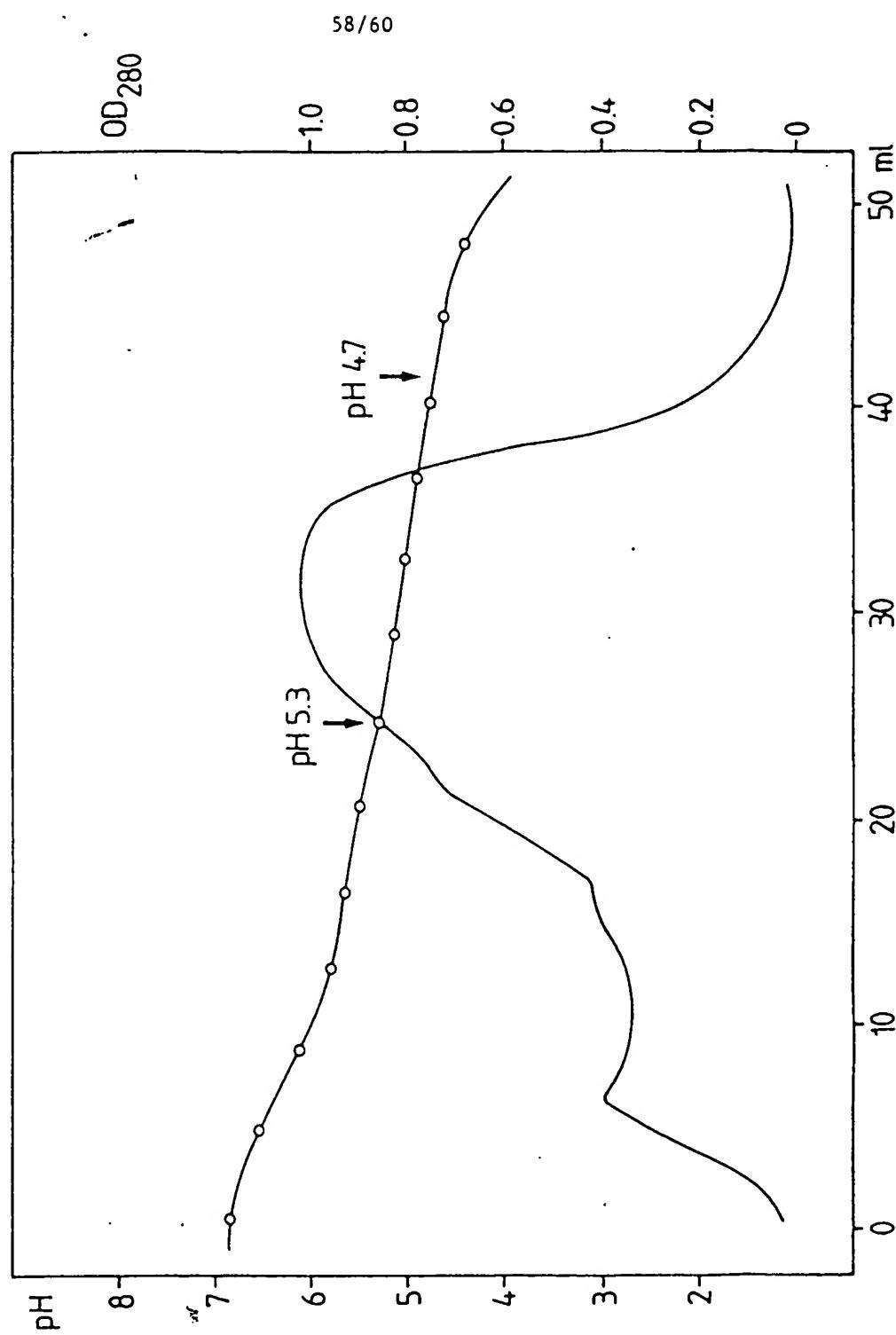


fig.14b

SUBSTITUTE SHEET

FIG. 15



SUBSTITUTE SHEET

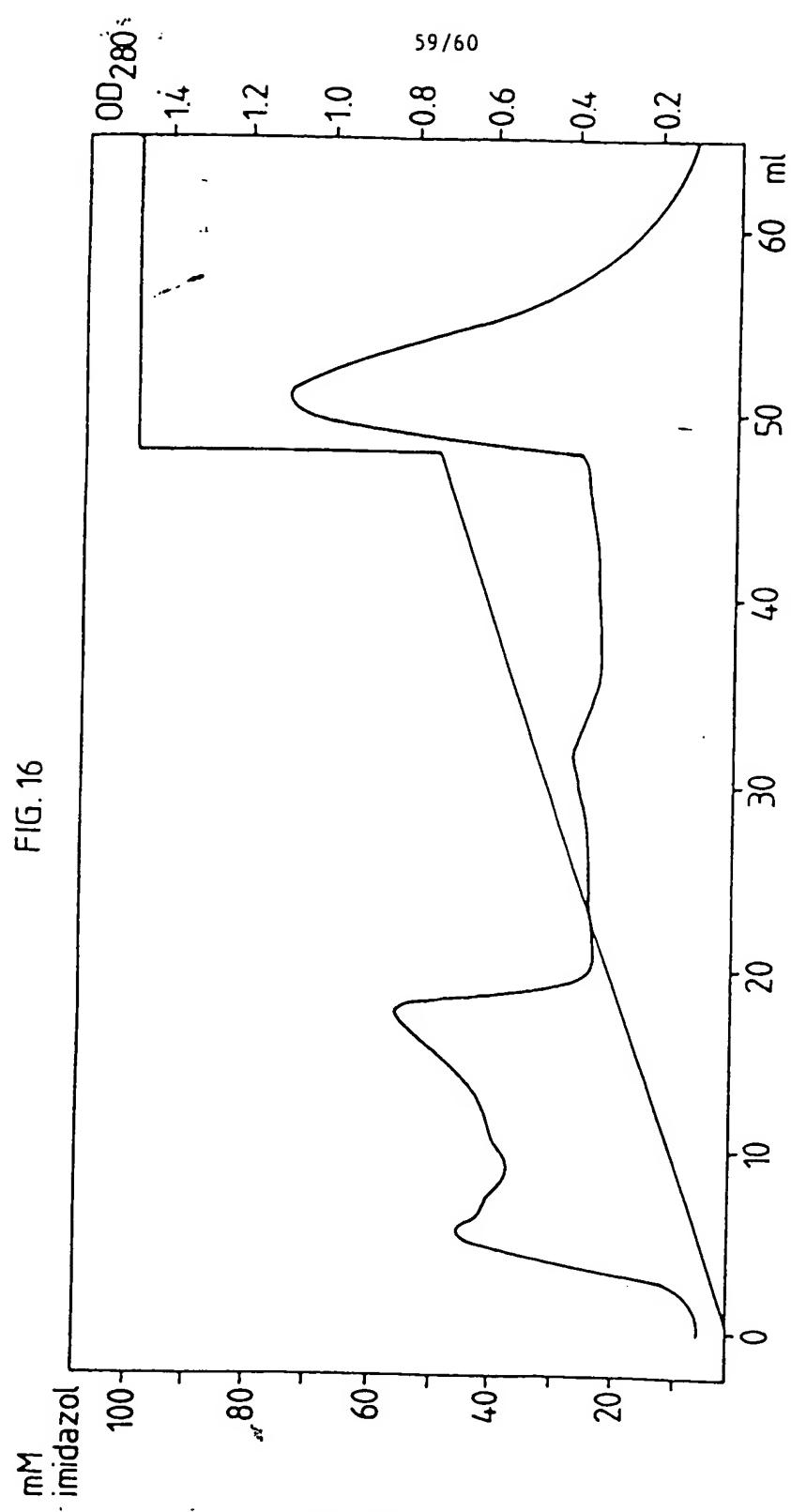
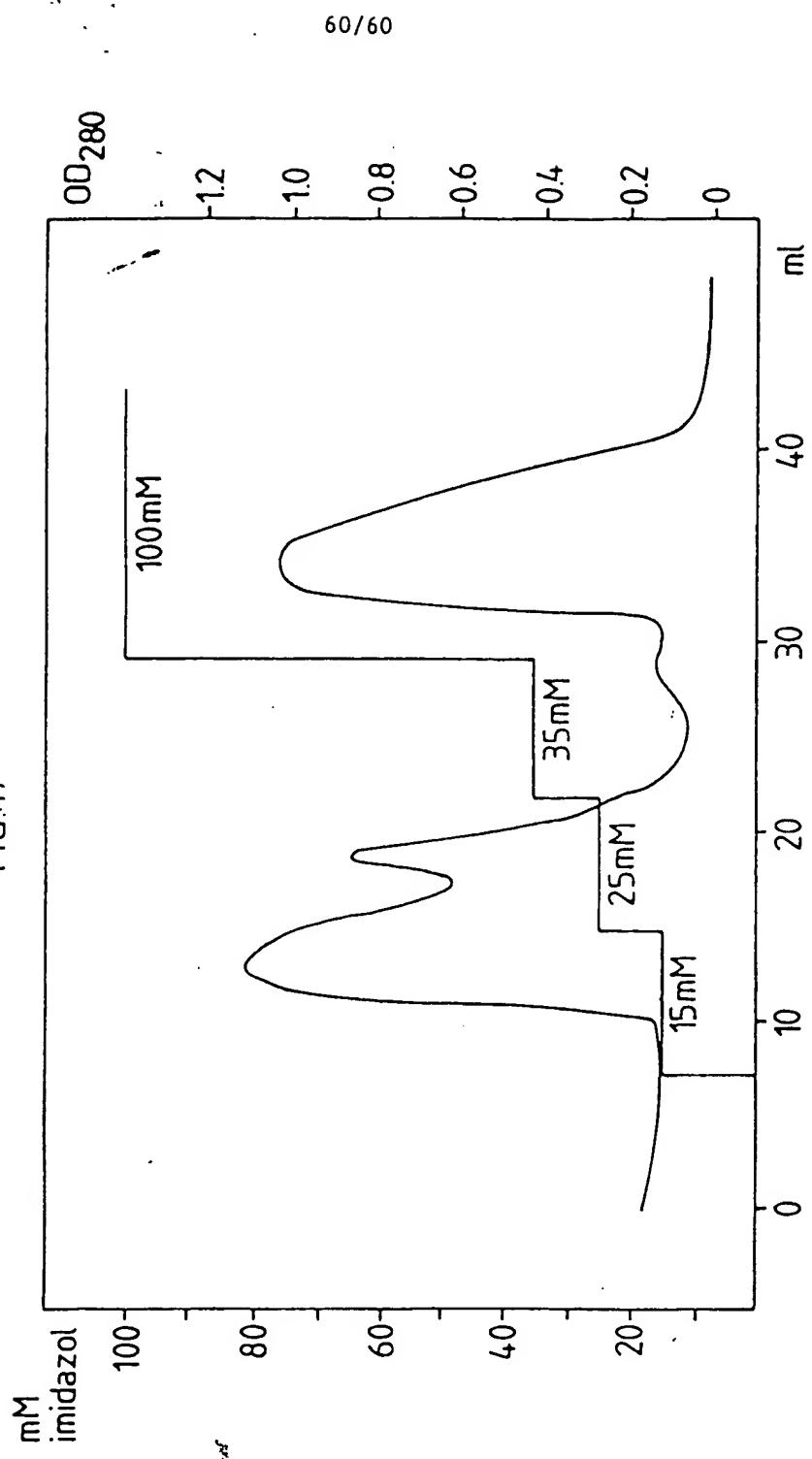


FIG.17



**SUBSTITUTE SHEET**

# INTERNATIONAL SEARCH REPORT

International Application No. PCT/EP 90/01593

## I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC: 5 C 07 K 13/00, A 61 K 39/04, C 12 N 15/31, G 01 N 33/569  
IPC: C 12 Q 1/68

## II. FIELDS SEARCHED

Minimum Documentation Searched \*

Classification System	Classification Symbols
IPC <sup>5</sup>	C 07 K, C 12 N, A 61 K, G 01 N, C 12 Q

Documentation Searched other than Minimum Documentation  
to the Extent that such Documents are Included in the Fields Searched \*

## III. DOCUMENTS CONSIDERED TO BE RELEVANT \*

Category \* | Citation of Document, \* with indication, where appropriate, of the relevant passages \*\* | Relevant to Claim No. \*\*\*

X	<p>Chemical Abstracts, volume 99, 12 September 1-9, 40, 41 1983, (Columbus, Ohio, US), H. Tasaka et al.: "Purification and antigenic specificity of alpha protein (Yoneda and Fukui) from mycobacterium tuberculosis and mycobacterium intracellulare", see page 413, abstract 86251m &amp; Hiroshima J. Med. Sci. 1983, 32(1), 1-8 (Eng). cited in the application</p> <p style="text-align: center;">---</p> <p>X   Journal of Clinical Microbiology, volume 25, no. 7, July 1987, M.L. Cohen et al.: " Expression of proteins of mycobacterium tuberculosis in escherichia coli and potential of recombinant genes and proteins for development of diagnostic reagents", page 1176 see the whole document cited in the application</p>	10-22, 25-33, 35-39, 43, 44

### --- Special categories of cited documents: \* "A" document defining the general state of the art which is not

considered to be of particular relevance

"E" earlier document but published on or after the international  
filing date

"L" document which may throw doubts on priority claim(s) or  
which is cited to establish the publication date of another  
citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or  
other means

"P" document published prior to the international filing date but  
later than the priority date claimed

"T" later document published after the international filing date  
or priority date and not in conflict with the application but  
cited to understand the principle or theory underlying the  
invention

"X" document of particular relevance; the claimed invention  
cannot be considered novel or cannot be considered to  
involve an inventive step

"Y" document of particular relevance; the claimed invention  
cannot be considered to involve an inventive step when the  
document is combined with one or more other such documents,  
such combination being obvious to a person skilled  
in the art.

"Z" document member of the same patent family

## IV. CERTIFICATION

Date of the Actual Completion of the International Search

20th December 1990

Date of Mailing of this International Search Report

23 JAN 1991

International Searching Authority

EUROPEAN PATENT OFFICE

Signature of Authorized Officer

Mme N. KUIPER

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**